Preprocessing Steps:

Why is text preprocessing done in NLP?

What are the advantages of using the Bag-of-Words (BoW) model to extract features?

How do you lowercase text in Python?

How do you remove punctuation from a text document?

What is tokenization, and why is it important?

Explain the difference between stemming and lemmatization.

How can you remove stop words from a text corpus?

What is the purpose of removing special characters and digits from text data?

How do you handle misspelled words during text preprocessing?

What is the role of part-of-speech tagging in NLP?

How can you handle synonyms and antonyms in text data?

What is the significance of removing HTML tags from web text?

How do you handle contractions (e.g., “don’t,” “can’t”) during preprocessing?

Explain the concept of n-grams and their use in feature extraction.

What is TF-IDF (Term Frequency-Inverse Document Frequency)?

How do you handle rare or infrequent words in a text corpus?

What is the purpose of removing URLs and email addresses from text data?

How can you normalize numerical values (e.g., dates, currency) in text?

What are regular expressions, and how are they used in text preprocessing?

How do you handle emoticons and emojis in text analysis?

Explain the concept of word embeddings (e.g., Word2Vec, GloVe).

How can you handle noisy or irrelevant text data?

What preprocessing steps are specific to social media text (e.g., tweets)?

How do you deal with imbalanced classes in text classification?

What techniques can improve the efficiency of text preprocessing for large datasets?

What is the purpose of stemming in text preprocessing, and how does it differ from lemmatization?

How can you handle misspelled words during text preprocessing?

Explain the concept of n-grams and their use in feature extraction.

What is the significance of removing HTML tags from web text?

How do you handle contractions (e.g., “don’t,” “can’t”) during preprocessing?

What are regular expressions, and how are they used in text preprocessing?

How do you handle emoticons and emojis in text analysis?

Explain the concept of word embeddings (e.g., Word2Vec, GloVe).

How can you handle noisy or irrelevant text data?

What preprocessing steps are specific to social media text (e.g., tweets)?

How do you deal with imbalanced classes in text classification?

What techniques can improve the efficiency of text preprocessing for large datasets?

How do you handle rare or infrequent words in a text corpus?

What is the purpose of removing URLs and email addresses from text data?

How can you normalize numerical values (e.g., dates, currency) in text?

What are some common techniques for handling negations (e.g., “not good”) in sentiment analysis?

How do you handle multi-word expressions (e.g., “New York City”) during tokenization?

What is the role of part-of-speech tagging in NLP?

How can you identify and remove duplicate sentences from a text corpus?

Explain the concept of stop-word lists and their impact on text preprocessing.

How do you handle text data in languages other than English?

What are some challenges specific to handling non-English text?

How can you detect and correct spelling errors in a text document?

What are some strategies for handling out-of-vocabulary words in NLP models?

How do you preprocess text data for sequence-to-sequence tasks (e.g., machine translation)?

Preporcessing :

What are missing values and null values and how do they differ?

Type1 and type 2 errors

What are the common causes and types of missing values and null values in data?

How do you detect and count missing values and null values in a dataset?

How do you handle missing values and null values in a dataset?

What are the advantages and disadvantages of different methods of handling missing values and null values, such as deletion, imputation, substitution, etc.?

How do you choose the appropriate method of handling missing values and null values for your problem?

How do you evaluate the impact of handling missing values and null values on the quality and reliability of your data analysis and modeling?

How do you handle missing values and null values in categorical variables and numerical variables?

How do you handle missing values and null values in time series data and spatial data?

How do you handle missing values and null values in text data and image data?

What are outliers and how do they differ from anomalies and noise?

What are the common causes and types of outliers in data?

How do you detect and count outliers in a dataset?

How do you handle outliers in a dataset?

What are the advantages and disadvantages of different methods of handling outliers, such as deletion, transformation, clipping, etc.?

How do you choose the appropriate method of handling outliers for your problem?

How do you evaluate the impact of handling outliers on the quality and reliability of your data analysis and modeling?

How do you handle outliers in categorical variables and numerical variables?

How do you handle outliers in time series data and spatial data?

How do you handle outliers in text data and image data?

What is the central limit theorem and what are its assumptions and implications?

How do you apply the central limit theorem to sample data and population data?

How do you use the central limit theorem to estimate the mean and standard deviation of a population from a sample?

How do you use the central limit theorem to construct confidence intervals and hypothesis tests for a population parameter?

How do you use the central limit theorem to compare the means of two or more populations from samples?

How do you use the central limit theorem to perform ANOVA and regression analysis?

How do you use the central limit theorem to check the normality of a distribution and perform normality tests?

How do you use the central limit theorem to perform bootstrap resampling and Monte Carlo simulation?

How do you use the central limit theorem to perform power analysis and sample size calculation?

How do you use the central limit theorem to perform meta-analysis and systematic review?

What is one-hot encoding and what are its applications and advantages?

How do you perform one-hot encoding on categorical variables in a dataset?

What are the disadvantages and challenges of one-hot encoding and how do you overcome them?

How do you handle missing values and null values in one-hot encoding?

How do you handle ordinal variables and nominal variables in one-hot encoding?

How do you handle high cardinality and low frequency in one-hot encoding?

How do you handle multicollinearity and dimensionality in one-hot encoding?

How do you handle new and unseen categories in one-hot encoding?

How do you compare one-hot encoding with other encoding methods, such as label encoding, binary encoding, frequency encoding, etc.?

How do you implement one-hot encoding in Python or R using packages such as pandas, sklearn, or onehotencoder?

What is label encoding and what are its applications and advantages?

How do you perform label encoding on categorical variables in a dataset?

What are the disadvantages and challenges of label encoding and how do you overcome them?

How do you handle missing values and null values in label encoding?

How do you handle ordinal variables and nominal variables in label encoding?

How do you handle high cardinality and low frequency in label encoding?

How do you handle new and unseen categories in label encoding?

How do you compare label encoding with other encoding methods, such as one-hot encoding, binary encoding, frequency encoding, etc.?

How do you implement label encoding in Python or R using packages such as pandas, sklearn, or labelencoder?

What is text to vector and what are its applications and advantages?

How do you perform text to vector on text data in a dataset?

What are the disadvantages and challenges of text to vector and how do you overcome them?

How do you handle missing values and null values in text to vector?

How do you handle different languages and dialects in text to vector?

How do you handle different formats and styles in text to vector?

How do you handle different levels of granularity and specificity in text to vector?

How do you compare text to vector with other text representation methods, such as bag of words, term frequency-inverse document frequency (TF-IDF), word embeddings, etc.?

How do you implement text to vector in Python or R using packages such as nltk, spacy, gensim, or text2vec?

Evalution of classification models

So how you are going to evaluate the classification model

When to use the recall, precision when to use the F1 school which scenario

So what is a recall

what is your precision

what is F1 score

what is a Roc curve

what is a AUC

What is a confusion matrix

what is a typhoid error

what is a type 2 error

what is a hypothesis

what is accuracy

what is a log loss

Evaluation of the regression models

So by using the root mean square error

using the mean square error

using the mean absolute error

Relative absolute error

What is R2 square error?

What is adjusted r2 square error?

Linear Regression:

-------------------

How do I interpret the p-value in regression models?

Can you explain Type I and Type II errors in hypothesis testing?

How do I choose the sample size for a hypothesis test?

What is assumptions of the linear regression

if the model doesn't follow the assumptions.

How does the linear regression works.

Logistic Regression:

When to use LR , When to use SVM, RF , XGBoost, Decision Tree,

Assumption of LR

Eventhough it do classification , why it is called as logistic regression.

What are the advantages and disadvantages of logistic regression compared to other classifiers?

When to use the logistic regression like in which situation.

What is the loss of logistic regression.

How do u calculate the thresthold of LR – we use ROC curve?

What is the formula for LR

Where the regularization terms added to functins in order to avaid the overfitting

How do you deal with overfitting or underfitting in logistic regression?

How do you implement logistic regression in Python. What are some applications or use cases of logistic regression in real-world scenarios?

Why square loss is not used in logistic regression?

What is the main function used in logistic regression?

How do you implement multinomial logistic regression?

If Is prone to noise and overfitting?

If there are n mumber of multiclasses , how many LR need

What is the mathematical intuition for logistic regression?

What are some examples of logistic regression use?

ROC is a plot of ---

L1 regurarization

L2 regularization

Parameters of Logistic regression

What is the difference between logistic regression and linear regression?

**What are the analytical challenges during model development?**

**SVM---**

Different types of kernels used in the SVM

so why the data is converted into the higher domination l and their it is segregated in SVM

what is the loss function we used in the SVM

What is the vectors points

what is a marigin

What is a support vectors

So what is the main importance of the SVM model compared to other models when it is going to use.

How is SVM works?

What are the advantages and disadvantages of SVM compared to other classifiers?

How do you interpret the coefficients of a SVM model?

How do you handle categorical or missing features in SVM?

How do you evaluate the performance of a SVM model?

How do you deal with overfitting or underfitting in SVM?

How do you implement SVM in Python or R?

What are some applications or use cases of SVM in real-world scenarios?

What are the roles of support vectors and margins in SVM?

What are the different types of kernels in SVM and how do they affect the decision boundary?

How do you choose the best kernel and hyperparameters for SVM?

How does SVM handle the issue of class imbalance?

How does SVM handle multi-class classification problems?

How does SVM handle non-linearly separable data?

What are the differences between linear SVM, SVC, SVR, and SVRG?

How do you compare SVM with other machine learning algorithms such as logistic regression, decision trees, random forests, neural networks, etc.?

How do you explain the concept of kernel trick in SVM?

How do you visualize the decision boundary and support vectors of a SVM model?

How do you scale and normalize the data for SVM?

How do you use cross-validation and grid search to optimize SVM?

How do you debug and troubleshoot SVM models

Decision Tree

----

What is a information gain

what is a genie index

what is entropy

what is a decision tree? how the decision tree works

on which basis point it takes the column for inorder to split a tree

when to use the decision tree compared to the logistic regression random forest SVM

show is it going to walk on the high dimensional data or the low dimensional data

What are some advantages and disadvantages of decision trees?

How do you prevent or reduce overfitting in decision trees?

How do you measure the performance of a decision tree?

What are some criteria for choosing the best split at each node of a decision tree?

What are some popular algorithms for constructing decision trees and what are their differences?

How do you handle missing values and categorical variables in decision trees?

How do you prune a decision tree and what are the benefits of pruning?

What is a random forest and how is it related to decision trees?

What are some applications of decision trees in real-world problems?

Random Forest:

-------------------------------------------

What is a random forest. How it works

What are the parameters in decision tree and random forest

How to find the best parameters

What is a bootstrap technique

When to use RF.

**How do you tune the hyperparameters of Random Forest?**

**What are the advantages and disadvantages of Random Forest**

What is the difference between Random Forest and Gradient Boosting?

What is the difference between Random Forest and Decision Trees?

Out of bag of words how to deal with it

What are the advantages and disadvantages of random forest over a single decision tree?

How does random forest handle missing values and outliers in the data?

How do you tune the hyperparameters of random forest, such as the number of trees, the maximum depth, the minimum samples per leaf, and the feature selection method?

How do you measure the importance of features in random forest?

How do you deal with imbalanced classes in random forest?

How do you evaluate the performance of random forest on a test set?

How do you prevent overfitting and underfitting in random forest?

How does random forest handle categorical and numerical features?

How do you compare random forest with other ensemble methods, such as boosting and stacking?

How do you implement random forest in Python or R?

Random forest is the extension of Bagging technique

What is Out-of-bag- error

What is the difference between batch gradient descent, stochastic gradient descent, and mini-batch gradient descent?

What are the advantages and disadvantages of stochastic gradient descent over batch gradient descent?

How do you choose the learning rate and the batch size for mini-batch gradient descent?

What are some techniques to improve the convergence and stability of stochastic gradient descent, such as momentum, Nesterov accelerated gradient, AdaGrad, RMSProp, and Adam?

How do you implement stochastic gradient descent in Python or R?

Gradient Boosting model:

----------------

What is Gradient boosing and how it works?

What are the main components of the gradient boosting algorithm and how do they interact with each other?

How does gradient boosting handle regression and classification problems differently?

What are some advantages and disadvantages of gradient boosting compared to other ensemble methods, such as bagging and stacking?

How do you tune the hyperparameters of gradient boosting, such as the learning rate, the number of trees, the tree depth, and the regularization parameters?

How do you measure the feature importance and the prediction error of gradient boosting?

How do you deal with missing values, outliers, and imbalanced classes in gradient boosting?

How do you compare and contrast gradient boosting with other popular boosting algorithms, such as AdaBoost and XGBoost?

How do you implement gradient boosting in Python or R using libraries such as scikit-learn, lightgbm, or catboost?

What is the difference between gradient boosting and random forest?

XGBoost:

---------------------

Where it is used the xgboost and how it works

What are the paramers we need to give to the XGboost

How does XGBoost use regularization to prevent overfitting? What are the two types of regularization terms that XGBoost supports?

How can you use XGBoost to perform feature selection? What are the two methods that XGBoost provides to rank the features by importance?

How can you tune the hyperparameters of XGBoost using cross-validation? What are some of the common hyperparameters that you need to optimize and how do they affect the model performance?

How can you use XGBoost for multi-class classification problems? What are the two options that XGBoost offers for the objective function and how do they differ?

How can you use XGBoost for imbalanced data sets? What are some of the techniques that XGBoost implements to handle class imbalance and improve the model accuracy?

How can you use XGBoost for time series analysis? What are some of the challenges and solutions that XGBoost offers for dealing with temporal data?

How can you use XGBoost for natural language processing tasks? What are some of the applications and benefits of using XGBoost for text data?

How can you use XGBoost for computer vision tasks? What are some of the advantages and disadvantages of using XGBoost for image data?

How can you use XGBoost for recommender systems? What are some of the ways that XGBoost can incorporate user and item features and interactions to generate personalized recommendations?

How can you use XGBoost for anomaly detection? What are some of the methods and metrics that XGBoost can use to identify outliers and anomalies in the data?

How does XGBoost use linear models as base learners instead of trees? What are the pros and cons of using linear models and how can you switch between them?

KNN:

---------

What is the KNN algorithm and how does it work?

What are the advantages and disadvantages of using the KNN algorithm?

How do you choose the optimal value of K in the KNN algorithm?

What are some of the distance metrics that can be used in the KNN algorithm?

How do you handle missing values and categorical variables in the KNN algorithm?

How do you evaluate the performance of the KNN algorithm?

How do you improve the efficiency and scalability of the KNN algorithm?

What are some of the applications of the KNN algorithm in real-world problems?

How do you compare the KNN algorithm with other machine learning algorithms such as linear regression, logistic regression, decision trees, etc.?

How do you implement the KNN algorithm in Python or R?

Naïve Bayes theorm--:

Assumption

What is a naïve bayes theorm

What is about the probality usings

What is the Naive Bayes algorithm and what is the basic assumption behind it?

What are the advantages and disadvantages of using the Naive Bayes algorithm?

How do you choose the appropriate probability distribution for the features in the Naive Bayes algorithm?

How do you handle missing values and categorical variables in the Naive Bayes algorithm?

How do you evaluate the performance of the Naive Bayes algorithm?

How do you compare the Naive Bayes algorithm with other machine learning algorithms such as logistic regression, decision trees, k-nearest neighbors, etc.?

How do you implement the Naive Bayes algorithm in Python or R?

What are some of the applications of the Naive Bayes algorithm in real-world problems?

How do you deal with zero-frequency problem in the Naive Bayes algorithm?

How do you handle imbalanced data in the Naive Bayes algorithm?

KMeans:--

----

How ur going to select the K value

How to evalueate the Kmeans model.

What is the K-means algorithm and what is the basic idea behind it?

How does the K-means algorithm work and what are the steps involved?

What are the advantages and disadvantages of using the K-means algorithm?

How do you choose the optimal number of clusters (k) in the K-means algorithm?

What are some of the methods to measure the quality and validity of the clusters obtained by the K-means algorithm?

How do you handle outliers, missing values, and categorical variables in the K-means algorithm?

What are some of the challenges and limitations of the K-means algorithm?

How do you compare the K-means algorithm with other clustering algorithms such as hierarchical clustering, DBSCAN, etc.?

How do you implement the K-means algorithm in Python or R?

What are some of the applications and use cases of the K-means algorithm in real-world problems?

LDA, DBSCAN, Topic Modelling:

-----

What is LDA and what are its applications?

How does LDA work and what are the main steps involved?

What are the assumptions and limitations of LDA?

How do you choose the number of topics and the number of words per topic in LDA?

How do you estimate the parameters of LDA using variational inference or Gibbs sampling?

How do you evaluate the quality and coherence of the topics generated by LDA?

How do you visualize and interpret the topics and the topic distributions of LDA?

How do you implement LDA in Python or R using packages such as gensim or topicmodels?

What are the differences and similarities between LDA and PCA?

What are the extensions and variations of LDA such as supervised LDA, correlated LDA, dynamic LDA, etc.?

What is DBSCAN and what are its applications?

How does DBSCAN work and what are the main steps involved?

What are the advantages and disadvantages of DBSCAN over other clustering algorithms?

How do you choose the parameters of DBSCAN such as epsilon and minPts?

How do you handle noise, outliers, and border points in DBSCAN?

How do you evaluate the performance and validity of the clusters obtained by DBSCAN?

How do you visualize and interpret the clusters and the cluster labels of DBSCAN?

How do you implement DBSCAN in Python or R using packages such as scikit-learn or dbscan?

What are the differences and similarities between DBSCAN and K-means?

What are the extensions and variations of DBSCAN such as OPTICS, HDBSCAN, etc.?

What is topic modeling and what are its applications?

How does topic modeling work and what are the main steps involved?

What are the advantages and disadvantages of topic modeling over other text analysis methods?

How do you choose the appropriate topic modeling technique for your problem?

How do you preprocess and prepare the text data for topic modeling?

How do you estimate the optimal number of topics for topic modeling?

How do you evaluate the quality and coherence of the topics generated by topic modeling?

How do you visualize and interpret the topics and the topic distributions of topic modeling?

How do you compare and contrast different topic modeling techniques such as LDA, NMF, LSA, etc.?

How do you implement topic modeling in Python or R using packages such as gensim, sklearn, nltk, etc.?

Deep Learning:

--------------

What is deep learning and how is it different from machine learning?

What are the main components of a deep learning model?

What are some of the common types of deep learning architectures and their applications?

What are the advantages and disadvantages of using deep learning over other methods?

How do you choose the appropriate deep learning framework for your project?

What are some of the popular deep learning frameworks and libraries and how do they compare?

How do you preprocess and prepare the data for deep learning models?

How do you split the data into training, validation, and test sets and why is it important?

How do you handle imbalanced, noisy, or missing data in deep learning models?

How do you design and implement a custom deep learning layer or module?

How do you initialize the weights and biases of a deep learning model and what are the effects of different initialization methods?

How do you select the appropriate activation function for a deep learning model and what are the effects of different activation functions?

How do you select the appropriate loss function for a deep learning model and what are the effects of different loss functions?

How do you select the appropriate optimizer for a deep learning model and what are the effects of different optimizers?

How do you tune the hyperparameters of a deep learning model and what are some of the methods and tools for hyperparameter optimization?

How do you regularize a deep learning model and what are some of the techniques and methods for regularization?

How do you monitor and evaluate the performance of a deep learning model and what are some of the metrics and tools for evaluation?

How do you debug and troubleshoot a deep learning model and what are some of the common errors and issues that you might encounter?

How do you save and load a deep learning model and what are some of the formats and tools for model serialization?

How do you deploy and scale a deep learning model and what are some of the platforms and tools for model deployment?

How do you update and maintain a deep learning model and what are some of the challenges and best practices for model maintenance?

How do you ensure the quality and reliability of a deep learning model and what are some of the methods and tools for testing and validation?

How do you ensure the security and privacy of a deep learning model and what are some of the threats and solutions for model security and privacy?

How do you ensure the fairness and ethics of a deep learning model and what are some of the issues and guidelines for model fairness and ethics?

How do you explain and interpret the results of a deep learning model and what are some of the methods and tools for model explainability and interpretability?

How do you improve the efficiency and speed of a deep learning model and what are some of the techniques and methods for model optimization and acceleration?

How do you improve the accuracy and robustness of a deep learning model and what are some of the techniques and methods for model improvement and enhancement?

How do you transfer the knowledge and skills of a deep learning model and what are some of the techniques and methods for model transfer and adaptation?

How do you generate new and novel data and outputs from a deep learning model and what are some of the techniques and methods for model generation and synthesis?

How do you learn from multiple and diverse sources of data and information with a deep learning model and what are some of the techniques and methods for model integration and fusion?

SQL

Explain the what the output we are going to get when it's a left join, right join, inner join and full outer join.

Try to get the outer joins this all the things when when we are having the null values also in the columns.

Applying the weather condition in the having condition

so applying the group by condition on the 2 to 3 columns at a time getting the same call average this kind of the information

what is a natural join

what is the difference between the delete, truncate, drop

So how to get the second highest salary after particular department.

It's like using getting the rank dense rank and the roll number of a window function.

So using the over( partistion by ordered by and that tense?

Joining of 2 tables for getting knowing information of foreign key and primary key

Getting using getting the errors of the different kinds of the values when they hit any API like 402 100 error 4 not 1581 error

**Python**

Difference between generator and iterator?

What is a decorator and give me an example of that one?

What is a map, reduce, filter functions in Python?

What is the pandas and numpy?

Do you know the operations using the loc in the iloc ?

Difference between the taple and list?

Program on the weather all the brackets are closed in a given expression or not you need to check how you are going to check that one?

Difference between the append and extend?

Get the number of the lines, characters, words In a given file?

What is the difference between multithreading and multiprocessing?

What is a list comprehension?

And write the Fibonacci series prime number programmes?

What is the\_double or the scroll name name equal to main?

What is the difference between the append and extend?

What is the use of init function or method?

How to handle the exception handling in Python? And try to get the even numbers from a list of elements?

What is a Lambda function in Python what is the main use of that function?

What is the difference between the equal to and ‘is’ in Python?

How the memory management is done in a Python?

What is a pythons global interpreter lock?

Which module you are going to use for the log to get the logs in Python?

5 Python is more famous what is the main reason?

Sorting of the dictionary programme?

Creation of a data frame?

Addition of 2 list and the addition has to be one is to one of the first element of the first element of the second list?

Use of get function in dictionary?

Sorting of the elements in a list?

Multithreading concept in Python?

How to call the parent class function from the child class in multithreading?

What is a method overloading and method overriding in Python?

What is encapsulation?

What is a Singleton class what is empty class?

What is a virtual method and abstract method?

Create a data frame by using the dictionary which does not contain any list items there you need to use the square brackets in the data frame paranthesis?

Create a 3 by 3 array by using the numpy?

So get the common substring from the all the list of items in suppose in a list cars carpet karpur carrom Carpenter is there you need to get the output as a car because it is a largest substring from all the items type equal to car Hey?

Show get the use of the Lambda function?

Lock in the eye lock difference?

Combining 2 data frames by using the merge join so when to use the join when to use the merge or when to use the catcat you must know the difference?

Reverse of a string?

What is a deep copy and shallow copy?

Try to reverse office string without using the column methods?

Perform the multiplication to a list of items perform the addition to the list of items?

What is a polymorphism?

What is encapsulation?

Performing the different type of operations on the data frame by using the pandas one?

Creation of arrays by using the numpy and performing the operations?

Why everything in a Python is an object?

Why Python is a functional scripting and object oriented language?

So write a program using the regular expression removing the numbers from a given string?

Raiche Python programme by passing the 2 numbers in the operations it need to be performed the addition subtraction multiplication and the division?

Use of the random module? And get the list of the numbers from zero to 100 ,10 numbers

Use of if for while?

Use of separator in a print function?

What is the main limitation of the Python?

What are different identifiers in a Python?

How the function in a Python is represented?

Try to learn about the naming conventions in Python?

Starts with a single under scroll and the double under scroll and double under scroll and ends with the double under scroll what does it represents? Python language defined identifiers?

How many reserved keywords are there in Python?

Sorting of the different elements in a list?

Ture as 1 and False as 0? Use of type keyword?

Try to see the binary octal hexadecimal representation of a number in a Python?

What is a star ark send the double star arguments passing as a parameter in a function?

How the function can pass a another function in a Python that is the concept of the decorator?

List[begin:end:step]

Type casting in Python?

What are different type of errors in Python?

So why immutability concept was introduced in Python?

All fundamental data types are immutable like int float string complex boolean are immutable?

Why in Python zero to 256 only change the common object for a int variable?

Different characters of a list that is related to the order heterogeneous remove element?

Range is a data type and range is a function both are available and it is immutable?

Use of the enumeration function in Python?

So try to get the different operations of the list tuple set dictionary?

Like remove and insert append extend C get in dictionary?

, dict {} - for dict

Set() --represe Representation?

Duplicate keys are not allowed in dictionary?

Even though the lists values are same doesn't mean that they both are referring to the same address in order to check the weather the 2 elements of the 2 data of 2 ‘==’.

a, b, c, d = 1,2,3,4

split(), join(), in print(sep=’’, end=’’)

replace operator = {} {0},{1}.format(name, sala, gf)

need to go with the for loop and go for the while(not know iteration untilt true condition) loop?

Pattern trees programs?

Break, continue and pass.

While else, for else,

S=’durga’

Del s, del and None – mult object if del obj not deleted, input()

Rstrip(), strip(),

find(), index() not availe value error comes in index() not there -1 comes.

Count() – substring in a given string

s.replace(), upper, lower, swapcase(), title() – EVERY wod uppercase, eemaiing lower, capatalize(), startswith, endswith()

isalnum, isalpha,isdigit, islower, isupper, istitle()

method when in class

function without class

insert()index,element

Difference between equal operator and copy operator in lis?x=y –alising, copy – cloing purpose

a+b it return to c, a.extend(b) it wont return with in exting

list + number ---error typeerror

list\*2 ---> two times the list elements will get.

List comprehension = [x\*x for I in range(0,99)]

N1[10,20,30, 40], n2=[40,50,60] present in n1 not in n2

Max, min utilization check it once

Count(), index(), sorted

t.sort() , sorted() – t.sort is not possible

a, b, c, d = tuple(t) – value error when try to change.

Difference between list and set, dict, tuple?

Different types of the parameters used in the functions like the default ,key word , variable length parameter positional parameters

Add(a,b), add(a,b, c) for this reason we went variable length and variable keyword paramers

In order to loop the kwg we need to use the for loop.

Tell me about the global keyword.

Get factorial of number?

Filter(func or lambda x:x%2==0, list)

Map(lambda l1,l2:x\*y, l1,l2)

S = lambda a,b:a+b ; s(3,4)

Function alisasing

Nested functions

Use of math module, random module, random function bet 0<x<1

Randomint(1,100), randrange(0,100,10)

Zerodivisonerror.ArthimaticeError, filenotfounderror,

Single except block can handle multiple explections.value error, invalidcouponcodeexcpe

Heart related but eye specialist so not able to handle.

Os.\_exit(0) – automatically vm shutdown.

Try -exception, else block, nested try block

How to create the custom exceptions.. if conditions and raise conditions used

Import logging module uses info, debug, defect..

Assert uses instead of print

Types of file handling variables like r, w , a, read(), readlines()

Os.path.isfile(‘dfdf’)

Operations on directories –cureent , new, existing, list of contents,

Os.getcwd(), os.mkrdirs(‘siva/python/db’), os.rmdir(), os.removedirs, os.rename

Listdir(), walk() it is a

Dirpath, dirnames, filesnames using walk function

Os.system()

Pickling and unpickling.

State of object to file

Obj contains 1000 properties to file ---> picking – other lang serialization

Dump

Load(file), why save object to the file –in future

Static, non statice variable

Import gc – garbaagee collector

Has and is a relationship in multithreading

In java constructoe not available to the child but in python everything Is available, exept iparent nstance variable by using self variable.

Same child and parent variable – then most recent variable.

Super key word- by using access parent everything

Super().\_\_init\_\_()

Regular expressions – re.compile to find the pattern /s,-space

/S except space

/d – any digit

/D – except dogits

\w= any word charc(alpha numerice

\W – any character except word) ---match at the beginning of the obhect

~,^

Findall,

MRO

Import threading

Only operater overload not other overloading

Polymernsims- loading, ridding

Generative AI – Project – RAG

What is a transformer

what is the difference between the chart gpt and bERT

what is the main importance of the BERT

what is encoder and what is a decoder

what is the self attention using mechanism in Transformers

what is encoding what is embedding

Why need in addition to the self attention mechanism the decoder utilises the cross attention mechanism to focus the relevant parts of the input sequence

could you please explain the architecture of bert model.

What is a positional encoding

what is a layer normalization

what is a positional encodings

one of the important thing is like self attention and the positional encodings.

Britain and the fine tuning transformer based model like the bird are typically pre trained on the Indian supervisor supervisor which allows them to learn rich presentation of this preteen models can be fine tune to specify download visual way reading to the improve the faster the transfer learning

what is a transfer learning

What is the drawback of the recursive neural network

what is meant by the embedding interview questions on the board model.

BERT Model.

What is the full form of the Bert model.

Can you please explain how Bert model is pre trained and fine tune for downstream task hey

what are the key components of the BERT architecture

how does the Bert handle the bidirectional contest in the natural language understanding task

what are some advantages of using the Bert compared to the traditional language models like the word embedding or the recurrent neural networks

can you describe the free train objectives used in the training of the be active model

how does the BET achieve the contextual representation of the words and tokens in a sentence

what are the some common applications of the Bert in the natural language processing beyond the text classification.

How does the BT model handle the out of vocabulary boards or the tokens

what are the some limitations or challenges associated with using the Bert model in real time world application

how can be it be fine tune for specific domain on specific task for specific nlp task

can you discuss any recent advancements or the variation of BRT in an NLP research community

So lot of tables are there so while using the joint it is taking a lot of time how we need to do the optimization.

Memphsis-

So the first question he had asked me regarding the at the distinct dates Richard the user had bought the product so try to give me that output so by using the one table he had given me in the chat window pause so I had done the group by of the users and the dead and taken the having the value of the count of the date is greater than one in that way I had failed that one.

And apart from the the second thing is like he had asked me the events occurred by the red and the e yellow events are there so I want the users who add take a wire cut the Roly the ready events I don't want the yellow event users that is another approach

and the 3rd thing is like he had asked me about the difference between the gradient descent stochastic and the mini batch stochastic gradient descent.

And different and simple techniques used it different ensemble technique.

So will you go with the accuracy model or interpretive model I told that most of the time the data is imbalanced dataset so I will go with the interpretation of the model.

How do you do the feature selection technique in the nlp natural language processing.

What are the different feature selection techniques are you today Indian LP nlp. Hey

Missing values normalization outliers imbalanced data set next is the feature selection

How to find the important features in a given data set?

Dropping the constant values.

By using the correlation function.

ChiSquare test.

So then never there is a lot of features are there means there will be a situation called as a curse of dimensionality will occurs. As the number of the features increases the accuracy decreases so in order to increase the accuracy we need the data at the 2 number of the features.

First we try to remove the features which is having the constant values throughout all the rows so for that we use the SK loan dot feature selection import the varianceThrusthold. Feature selector removes all the low variant features. Variance = 0.

After splitting we do the. Fit or the transform because we avoid the overfitting which is done told by the Krishna Naik. I need to look into this one.

So another one thing is like if they invent features are more highly correlated with the dependent features than it is a highly important feature. So we are not going to remove that one, it plays a prominent role.

But Not within the. Independent features. If they are highly correlated within the each other than it is better to take the. One feature. Suppose if there are 3 features here there out of which WI Fi. Take the one feature is enough. Show here we are taking the correlation of the data set and we are giving the. Trust hold for the correlation matrix and we are removing the unwanted. Features.

Mutal\_info\_classifier.

Show there are so many various techniques are there in this random forest exit boost recursive feature elimination and the correlation analysis these are the 4 methods I know very well by this we can get the important features and also lasso regression technique is the one of the technique by this we can get the important features.

In random forest the model thought fit XY means we will get the model by using that model model dot feature get feature importance we are going to get the feature importance why this we are going to get how often that feature is used in the split of the data in this way we use the random forest to get the important features.

The same way the exhibit boost also having the exist exit classified or 5th of xy so there also feature importance is the one of the function is there it is going to give the what are the important features similar to the random forest these models.

By using the recursive feature elimination we can get the important features also here suppose if we give the if we take the logistic regression then from the SQL if we take the feature selection and import r fee we are going to take the logistic regression model and in rfe if we take the parameter as a model and the number of the features to be selected so the 50 and we are going to get the rank of the each features.

So and also we are having the skull features the select the K best to select the top K based statistical test on statistical test.

Lasso is the one of the technique we use it escalator linear model where the last so we can import it so the lasso and if we give the alpha that is the learning rate and model dot fit XXY NP dots some modal questions medical to zero non zero positions so these are the methods we can use it in order to get the top and features are to get the important features.

Cognigent ---

My Project is total draw back for me

He asked me about

Wipro -the feedback is good for the Wipro he told that you are knowing about the generate way its adds a flash to your profile apart from that try to load the deep flooring gods of channel kick off and very carefully it adds some more ads to the profile. Hey

Introduction of myself to that 1I need to improve a lot.

And coming to the projects he has asked me about the nltk and spacy

what is the tokenizer you had used it?

And lemmatization and the stemming difference

what is embedding in different kinds of the embedding used it,

how the vo to wrd2ector works and how you are going to bill that one

difference between the Elysium and the recurrent neural network

tell me about the project related to the deep learnings which you had used it

so an opinion you had used the text embedding one model and the class model you had used it I think?

----how do u create the wokspace

GEN AI:

What is RAG and what are its main components and objectives?

How does RAG combine the capabilities of a large language model (LLM) and an external data source?

What are the advantages and disadvantages of using RAG over other methods of text generation?

How do you choose the appropriate data source and retrieval mechanism for RAG?

How do you evaluate the quality and relevance of the retrieved data for RAG?

How do you integrate the retrieved data into the LLM prompt for RAG?

How do you evaluate the quality and coherence of the generated text by RAG?

How do you handle missing values, null values, and outliers in the retrieved data for RAG?

How do you handle different languages, formats, and styles in the retrieved data for RAG?

How do you handle different levels of granularity and specificity in the retrieved data for RAG?

What are some of the common use cases and applications of RAG?

What are some of the challenges and limitations of RAG?

How do you optimize the performance and efficiency of RAG?

How do you ensure the security and privacy of the data used by RAG?

How do you ensure the fairness and ethics of the text generated by RAG?

How do you explain and interpret the results of RAG?

How do you update and maintain the data source and the LLM for RAG?

How do you deploy and scale RAG for production?

How do you compare RAG with other text generation techniques, such as GPT-3, GPT-4, etc.?

How do you implement RAG in Python or R using packages such as Hugging Face Transformers, PyTorch, or TensorFlow?

What are some of the current research topics and trends in RAG?

What are some of the open problems and challenges in RAG?

What are some of the best practices and tips for using RAG?

What are some of the resources and tools for learning and developing RAG?

How do you test and debug RAG and what are some of the common errors and issues that you might encounter?

How do you collaborate and communicate with other developers and stakeholders on RAG projects?

How do you document and report your RAG work and results?

How do you design and conduct experiments and evaluations for RAG?

How do you review and critique RAG papers and publications?

How do you keep up with the latest developments and innovations in RAG?