#coding Questions

==================

1 Write a NumPy program to find the set difference between two arrays. The set difference will return sorted, distinct values in array1 that are not in array2.

Expected Output:

Array1: [ 0 10 20 40 60 80]

Array2: [10, 30, 40, 50, 70, 90]

Set difference between two arrays:

[ 0 20 60 80]

Ans :-

import numpy as np

array1 = np.array([0, 10, 20, 40, 60, 80])

array2 = np.array([10, 30, 40, 50, 70, 90])

result = np.setdiff1d(array1, array2)

print("Array1:", array1)

print("Array2:", array2)

print("Set difference between two arrays:")

print(result)

2 Write a NumPy program to construct an array by repeating.

Sample array: [1, 2, 3, 4]

Expected Output:

Original array

[1, 2, 3, 4]

Repeating 2 times

[1 2 3 4 1 2 3 4]

Repeating 3 times

[1 2 3 4 1 2 3 4 1 2 3 4]

Ans:-

import numpy as np

original\_array = np.array([1, 2, 3, 4])

print("Original array")

print(original\_array)

result\_2\_times = np.tile(original\_array, 2)

print("Repeating 2 times")

print(result\_2\_times)

result\_3\_times = np.tile(original\_array, 3)

print("Repeating 3 times")

print(result\_3\_times)

3 Load diamond dataset and perform EDA.

Data=pd.read\_csv(‘filename.csv’)

4 Count the Number of matching characters in a pair of string.

Input : str1 = 'abcdef'

str2 = 'defghia'

Output : 4

(i.e. matching characters :- a, d, e, f)

Input : str1 = 'aabcddekll12@'

str2 = 'bb22ll@55k'

Ans:

def count\_matching\_characters(str1, str2):

common\_characters = set(str1) & set(str2)

count = len(common\_characters)

return count

str1\_1 = 'abcdef'

str2\_1 = 'defghia'

output\_1 = count\_matching\_characters(str1\_1, str2\_1)

print(f"Output 1: {output\_1}")

5 Python program for removing i-th character from a string

Input : Peter

i = 4

Output : Pete

input\_str = "Peter"

my\_list = list(input\_str)

# Remove element at index 4

index\_to\_remove = 4

removed\_element = my\_list.pop(index\_to\_remove)

print("List after removing element at index", index\_to\_remove, ":", my\_list)

result\_string = ' '.join(removed\_element)

Theory Question

1 Mention the different types of Data Structures in Pandas?

Series:

A one-dimensional labeled array that can hold any data type.

It is similar to a column in a spreadsheet or a single column in a DataFrame.

DataFrame:

A two-dimensional labeled data structure with columns that can be of different types.

Index:

An immutable array used for indexing and aligning data in Pandas.

It is similar to the row labels in a DataFrame or the axis labels in a Series.

Panel (deprecated):

A three-dimensional data structure that was used for handling 3D data in older versions of Pandas.

2:Explain SVM Algorithm in Detail

support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. It is particularly effective in high-dimensional spaces and is well-suited for tasks where there is a clear margin of separation between classes. The main idea behind SVM is to find the hyperplane that best separates different classes in the feature space.

3:Explain the Difference Between Classification and Regression?

| ****Classification**** | ****Regression**** |
| --- | --- |
| **In this problem statement, the target variables are discrete.** | **In this problem statement, the target variables are continuous.** |
| **Problems like [Spam Email Classification](https://www.geeksforgeeks.org/detecting-spam-emails-using-tensorflow-in-python/), [Disease prediction](https://www.geeksforgeeks.org/disease-prediction-using-machine-learning/) like problems are solved using Classification Algorithms.** | **Problems like [House Price Prediction](https://www.geeksforgeeks.org/house-price-prediction-using-machine-learning-in-python/), [Rainfall Prediction](https://www.geeksforgeeks.org/ml-rainfall-prediction-using-linear-regression/) like problems are solved using regression Algorithms.** |
| **In this algorithm, we try to find the best possible decision boundary which can separate the two classes with the maximum possible separation.** | **In this algorithm, we try to find the best-fit line which can represent the overall trend in the data.** |
| **[Evaluation metrics](https://www.geeksforgeeks.org/metrics-for-machine-learning-model/) like Precision, Recall, and F1-Score are used here to evaluate the performance of the classification algorithms.** | **Evaluation metrics like [Mean Squared Error,](https://www.geeksforgeeks.org/python-mean-squared-error/) [R2-Score](https://www.geeksforgeeks.org/ml-r-squared-in-regression-analysis/), and  [MAPE](https://www.geeksforgeeks.org/how-to-calculate-mape-in-python/) are used here to evaluate the performance of the regression algorithms.** |

4:What do you understand by Decision Tree in Machine Learning?

Decision trees are a popular [machine learning algorithm](https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/" \t "https://www.analyticsvidhya.com/blog/2021/08/decision-tree-algorithm/_blank) that can be used for both regression and classification tasks. They are easy to understand, interpret, and implement, making them an ideal choice for beginners in the field of machine learning. In this comprehensive guide, we will cover all aspects of the decision tree algorithm, including the working principles, different types of decision trees, the process of building decision trees, and how to evaluate and optimize decision trees.

5:What are the common ways to handle missing data in a dataset?

**Method 1** is deleting rows or columns.  
We usually use this method when it comes to empty cells.  
For example, if the majority of our data is missing for a column or for a row, we can simply delete them.

**Method 2** is replacing the missing data with aggregated values.  
In this case, we can calculate the aggregated value based on the rest of the values we have in the column and put the received number to the empty spot

**Method 3** is creating an unknown category.  
Categorical features have a number of possible values, which gives us an opportunity to create one more category for the missing values. This way we will lower the variance by adding new information to the data. This could be used when the original information is missing or cannot be understood,

**Method 4** is predicting missing values.  
where we have no missing values, we can train a statistical or machine learning algorithm in order to predict the missing values. Since among the samples for which this training is performed, there are missing values, it is necessary to replace them initially using one of the simplest methods for recovering gaps.