

Assignment No.1

Zeenat Sultan

Csc-20f-194

Computer Science

6th-Sec(C)

Artificial Intelligence

Submitted to :

Ms. Aqsa Umar

Q.1. You are supposed to come up with the dataset of your choice [the taken dataset should be like itshould be publicly with available], two members [two datasets but with same topic].

Let's choose the Fashion MNIST dataset as our image dataset for analysis. It's a publicly available dataset that contains grayscale images of clothing items belonging to 10 different categories. We'll create two versions of this dataset, one using NumPy and another using pandas, to demonstrate basic analysis techniques.

**1. Fashion MNIST Dataset using NumPy**

import numpy as np

# Load Fashion MNIST dataset using NumPy

# Assume fashion\_mnist\_data.npy contains the Fashion MNIST dataset

fashion\_mnist\_data = np.load('fashion\_mnist\_data.npy', allow\_pickle=True)

# Basic analysis using NumPy

# Calculate the shape of the dataset

num\_samples, image\_width, image\_height = fashion\_mnist\_data.shape

print("Shape of Fashion MNIST dataset:", fashion\_mnist\_data.shape)

# Calculate the mean pixel intensity using NumPy

mean\_pixel\_intensity = np.mean(fashion\_mnist\_data)

print("Mean Pixel Intensity (NumPy):", mean\_pixel\_intensity)

# Generate questions based on the analysis

print("\nQuestions:")

print("1. How many samples are there in the Fashion MNIST dataset?")

print("2. What are the dimensions of each image in the dataset?")

print("3. What is the mean pixel intensity of the images?")

print("4. Can we visualize sample images from each category?")

**2.** Fashion MNIST Dataset using pandas

import pandas as pd

# Load Fashion MNIST dataset using pandas

# Assume fashion\_mnist.csv contains the Fashion MNIST dataset

fashion\_mnist\_df = pd.read\_csv('fashion\_mnist.csv')

# Basic analysis using pandas

# Display the shape of the dataset

print("Shape of Fashion MNIST dataset:", fashion\_mnist\_df.shape)

# Display the first few rows of the dataset

print("First few rows of Fashion MNIST dataset:")

print(fashion\_mnist\_df.head())

# Calculate summary statistics using pandas

summary\_statistics = fashion\_mnist\_df.describe()

print("\nSummary statistics of Fashion MNIST dataset:")

print(summary\_statistics)

# Generate questions based on the analysis

print("\nQuestions:")

print("1. How many samples are there in the Fashion MNIST dataset?")

print("2. What are the unique categories/classes present in the dataset?")

print("3. What are the summary statistics of pixel intensity values?")

print("4. How does the distribution of pixel intensity values vary across different classes?")

Q.2 Apply Numpy and pandas till today we have studied, if you think you lack any understanding of

them (refer to their documentation). Analysis the data and take out the questions from

them(Have given demo in the class)

NumPy and pandas to analyze a sample dataset. We'll start by loading some sample data, performing basic analysis using both libraries, and then generating questions based on the analysis.

use the CIFAR-10 dataset as our image dataset for analysis. We'll analyze it using NumPy and pandas, performing basic exploratory data analysis and generating questions based on the analysis.

import numpy as np

import pandas as pd

# Sample data: CIFAR-10 dataset

# For demonstration purposes, let's assume we have a NumPy array containing image data

# CIFAR-10 typically has dimensions (num\_samples, width, height, channels)

# Here, we'll generate random data as a placeholder

num\_samples = 5000

image\_width = 32

image\_height = 32

num\_channels = 3 # RGB channels

cifar\_data = np.random.randint(0, 256, size=(num\_samples, image\_width, image\_height, num\_channels), dtype=np.uint8)

# Convert the NumPy array to a pandas DataFrame

# Reshape the data to have one row per image

cifar\_df = pd.DataFrame(data=cifar\_data.reshape(num\_samples, -1))

# Basic analysis using NumPy and pandas

# Calculate mean pixel intensity using NumPy

mean\_pixel\_intensity = np.mean(cifar\_data)

print("Mean Pixel Intensity (NumPy):", mean\_pixel\_intensity)

# Calculate mean pixel intensity using pandas

mean\_pixel\_intensity\_pd = cifar\_df.mean().mean()

print("Mean Pixel Intensity (pandas):", mean\_pixel\_intensity\_pd)

# Generate questions based on the analysis

print("\nQuestions:")

print("1. What is the mean pixel intensity of the images in the CIFAR-10 dataset?")

print("2. How does the pixel intensity vary across different channels (R, G, B)?")

print("3. Can we visualize the distribution of pixel intensities using histograms?")

print("4. Are there any outlier images with unusually high or low pixel intensities?")

print("5. How does the mean pixel intensity compare to other descriptive statistics such as median or standard deviation?")

These questions are designed to explore various aspects of the CIFAR-10 dataset and guide further analysis using NumPy and pandas. They can help gain insights into the characteristics of the image data and identify potential patterns or anomalies.