Practical No. 3: Implement program to read dataset (Text, CSV, JSON , XML)

Files for practical You can download and use these:

Text: https://raw.githubusercontent.com/selva86/datasets/master/sample.txt

CSV: https://raw.githubusercontent.com/mwaskom/seaborn-data/master/iris.csv

JSON: https://raw.githubusercontent.com/typicode/demo/master/db.json

XML: https://www.w3schools.com/xml/simple.xml

IX. Conclusion

In this practical, we learned how to read datasets from different file formats such as Text, CSV, JSON, and XML using Python. Reading data is the first and most important step in machine learning. This practical helps us understand different data formats and how to load them for preprocessing. It is useful for handling real-world datasets used in machine learning applications.

X. Practical Related Questions – Answers

1. What is a dataset and why is it important in data analysis?

A dataset is a collection of data stored in files. It is important because:

Machine learning works on data

Data is used for analysis and prediction

Without data, machine learning is not possible

2. How do you read a .txt file using Python?

A .txt file is read using the open() function in read mode.

```
In [ ]:  file = open("sample.txt", "r", encoding="utf-8")
         print(file.read())
         file.close()
```

3. Write a Python program to read a .csv file and display the first 5 rows.

```
In [ ]:  import csv

         file = open("iris.csv", "r")
         reader = csv.reader(file)

         for i, row in enumerate(reader):
             if i < 5:
                 print(row)

         file.close()
```

4. Write a program to read an XML file and extract specific tag values.

```
In [ ]:  import xml.etree.ElementTree as ET

         tree = ET.parse("simple.xml")
         root = tree.getroot()

         for person in root:
             print("Name:", person.find("name").text)
             print("Price:", person.find("price").text)
```

5. What is the difference between JSON and XML in terms of structure and use?

JSON	XML
Uses key–value pairs	Uses tags
Easy and compact	Lengthy format
Fast to read	Slower to read
Used in APIs and ML	Used in documents

6. Write a program to read an XML file and extract specific tag values.

```
In [ ]:  import xml.etree.ElementTree as ET

         tree = ET.parse("simple.xml")
         root = tree.getroot()

         for food in root:
             print("Name:", food.find("name").text)
             print("Price:", food.find("price").text)
             print("Calories:", food.find("calories").text)
             print("----------------")
```

XI. Exercise

Python program to read and display data from four different files

```
In [ ]:  import csv
         import json
         import xml.etree.ElementTree as ET

         # Read Text file
         print("Text File:")
         print(open("sample.txt", "r", encoding="utf-8").read())
```

```
In [ ]:  # Read CSV file
         print("\nCSV File:")
         for row in csv.reader(open("iris.csv")):
             print(row)
```

```
In [ ]:  # Read JSON file
         print("\nJSON File:")
         print(json.load(open("db.json")))
```

```python
# Read XML files
tree = ET.parse("simple.xml")
root = tree.getroot()

for food in root.findall("food"):
    name = food.find("name").text
    price = food.find("price").text
    print(name, "-", price)
```