CS23334 -Fundamental Of Data Science

240701463

Analysis of Data Types

Exp :1C

Date: 29-07-2025

Aim:

To **differentiate Structured, Semi-structured, and Unstructured data** by creating small representative datasets, loading them using Python, and describing their characteristics.

Algorithm:

1. Create a **Structured** dataset (CSV format) with fixed schema (columns).
2. Create a **Semi-structured** dataset (JSON format) with tags and a flexible structure.
3. Create an **Unstructured** dataset (TXT format) containing free-form text.
4. Load and print the structured data using **Pandas**.
5. Load and print the semi-structured data using the **json** library.
6. Load and print the unstructured data as a simple string.

Code:

```
import pandas as pd

import json

# Load Structured Data (from structured_data.csv)

structured = pd.read_csv("structured_data.csv")

print("Structured Data:\n", structured, "\n")

# Load Semi-Structured Data (from semi_structured_data.json)

with open("semi_structured_data.json") as f:

    semi = json.load(f)
```

```
print("Semi-Structured Data:\n", semi, "\n")
# Load Unstructured Data (from unstructured_data.txt)
with open("unstructured_data.txt") as f:
    unstructured = f.read()
print("Unstructured Data:\n", unstructured)
```

Output:

Structured Data:

	ID	Name	Age	Salary
0	1	John	28	50000
1	2	Emma	31	62000
2	3	Liam	25	48000

Semi-Structured Data:

[{'Name': 'John', 'Skills': ['Python', 'SQL'], 'Experience': 3}, {'Name': 'Emma', 'Skills': ['R', 'Tableau'], 'Experience': 5}, {'Name': 'Liam', 'Skills': ['Python', 'Excel'], 'Experience': 2}]

Unstructured Data:

John is a data scientist with strong analytical skills.

Emma loves exploring datasets and building dashboards.

Liam is passionate about Python and data visualization.

Result:

The experiment successfully loaded and differentiated the three data types, demonstrating their distinct organization, complexity, and suitability for different analytical tools. Thus the python program was executed successfully, and the output is verified.