

Assignment-1

① Define data science and mention applications.

Ans Data Science is an interdisciplinary field that uses scientific methods, algorithms and systems to extract insight and knowledge from structured and unstructured data.

Applications →

- (i) Healthcare.
- (ii) Finance
- (iii) Marketing.
- (iv) Transportation.

② Explain the working module of matplotlib with python coding.

Ans Matplotlib is a plotting library for Python used to create static, interactive and animated visualization.

Example →

```
import matplotlib.pyplot as plt
```

```
x = [1, 2, 3, 4, 5]
```

```
y = [10, 20, 25, 40, 50]
```

```
plt.plot(x, y, label='line 1')
```

```
plt.title('Simple Plot')
```

```
plt.xlabel('X-axis')
```

```
plt.ylabel('Y-axis')
```

```
plt.legend()
```

```
plt.show()
```

③ What is delimited files? Explain with python code.

Ans Delimited files are text files where each line represents a record, and fields within each record are separated by a specific character, known as a delimiter. Common examples include CSV and TSV files.

Q
Nos.

Example →

```
import csv
```

```
data = [
    ['Name', 'Age', 'City'],
    ['XYZ', 18, 'Bangalore']
]
```

```
with open('people.csv', 'w', newline='') as file:
    writer = csv.writer(file)
    writer.writerows(data)
```

④ Explain the working of web scrapping with Python code.

Ans → Web scraping involves extracting data from websites using automated scripts. In Python, libraries like 'requests' and 'BeautifulSoup' are commonly used for this purpose.

Example →

```
import requests
```

```
from bs4 import BeautifulSoup
```

```
url = 'https://example.com'
```

```
response = requests.get(url)
```

```
soup = BeautifulSoup(response.content, 'html.parser')
```

```
titles = soup.find_all('h2')
```

```
for title in titles:
```

```
    print(title.get_text())
```


5) Explain the difference between stdin and stdout, and how they relate to standard input and standard output streams in python.

Ans → In Python, stdin and stdout are file-like objects used for input and output operations. stdin refers to the stream from which a program reads input data while stdout is the stream to which a program writes output data.

Example →

```
user_input = input("Enter something:")  
print("You entered:", user_input)
```


Assignment-2.

Q.1. Discuss the benefits of using APIs for accessing data, service or functionality in Python applications.

Ans. Using APIs in Python has various benefits →

(i) It provides a direct way to access data, services, functionalities without the need to scrape data or reinvent the wheel.

(ii) It offers real-time data, which is crucial for up-to-date information.

(iii) It allows applications to scale by offloading complex processes to external services.

(iv) It enables integration with various platforms.

Q.2. What is KNN? Explain its significance in ML.

Ans. The KNN model is simple, non-parametric, and instance-based learning algorithm used for classification and regression in ML.

→ It works by identifying K-nearest data points to a given query point and making predictions based on the majority class or average value.

→ The significance of KNN lies in its simplicity and effectiveness for small datasets, ease of implementation.

Q.3. With the example of the Iris Dataset how KNN is done.

Ans. `import numpy as np`
`import pandas as pd`

`from sklearn.datasets import load_iris`

`from sklearn.model_selection import train_test_split`

`from sklearn.neighbors import KNeighborsClassifier`

`from sklearn.metrics import accuracy_score`

Q
Nos.

```
iris = load_iris()
X = iris.data
y = iris.target
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.3, random_state=42)
```

```
Knn Knn = KNeighborsClassifier(n_neighbors=5)
```

```
Knn.fit(X_train, y_train)
```

```
y_pred = Knn.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy Score: {accuracy:.2f}')
```

(4) What is logistic regression?

Ans: It is a statistical method used for binary classification that models the probability of a binary outcome based on one or more predictor variables.

It uses the logistic function to map predicted values to probabilities between 0 and 1.

It is defined by $\sigma(x) = \frac{1}{1+e^{-x}}$, where x is the linear combination of the input features.

(5) Explain Support Vector Machine.

Ans: SVM is a supervised machine learning algorithm used for classification and regression tasks.

→ The primary goal of SVM is to find the optimal hyperplane that maximizes the margin between different classes in the feature space.

→ The hyperplane is chosen such that it separates the classes with the largest possible margin, which is defined as the distance b/w the hyperplane and the nearest data points from each class.

→ Key features →

(i) Maximal Margin Classifier

(ii) Kernel Trick

(iii) Robust to Outliers