

Machine Learning

Question1:

Explain what metacharacters are in regular expressions and provide examples of commonly used metacharacters. Discuss their significance in pattern matching?

Question2: (Email Validation for User Registration)

Many websites require users to register using their email addresses. Develop a Python script that validates email addresses using regular expressions. Consider various email formats and ensure that the script can accurately distinguish between valid and invalid email addresses.

Email DEMO data:

user@example.com
user123@gmail.com
user.name@domain.co.uk
user123@sub.domain.com
123user@domain.com
user123@123domain.com
user@localhost
user@invalid
user@.com
user@domain
user@example
user123@
invalidemail@.com
@domain.com
user@123.456
user123@domain..com

Question3: (User Log Analysis for Error Detection)

Analyze log files from a web server or application and extract relevant information using regular expressions. Develop a Python script that reads log files, identifies error messages based on predefined patterns, and generates a summary report with details of errors encountered.

Assignment 5

Sample Log Data:

2024-04-18 08:30:45 [INFO] Application started successfully.
2024-04-18 08:32:15 [ERROR] Database connection failed: Connection timed out.
2024-04-18 08:35:22 [ERROR] Failed to process request: Invalid input data.
2024-04-18 08:40:11 [WARNING] Disk space running low: 90% usage detected.
2024-04-18 08:45:03 [ERROR] Server crashed: Segmentation fault.
2024-04-18 08:48:55 [INFO] New user registered: username123.
2024-04-18 08:50:27 [ERROR] Authentication failed: Invalid credentials.
2024-04-18 08:55:09 [ERROR] Internal server error: Unable to connect to external service.
2024-04-18 09:00:04 [INFO] Application stopped.

This sample log data contains various log entries with different log levels (INFO, ERROR, WARNING) and corresponding messages. There are also different types of errors, such as database connection failures, authentication errors, disk space warnings, and internal server errors. Additionally, the log entries include timestamps for each event.

This data can be used to test the log analysis script's ability to correctly identify and extract error messages, classify errors by severity, and generate a summary report with details of errors encountered.

Question4: (Web Scraping and Information Retrieval)

Implement a web scraping tool that retrieves information from web pages based on specified patterns using regular expressions. Develop a Python script that accesses web pages, extracts desired content (e.g., product prices, article titles), and stores the extracted data for further analysis.

Sample web page content

```
<!DOCTYPE html>
<html>
<head>
  <title>Sample Web Page</title>
</head>
<body>
<h1>Latest Products</h1>
```

Assignment 5

```
<div class="product">
  <h2>Product 1</h2>
  <p>Price: $19.99</p>
</div>
```

```
<div class="product">
  <h2>Product 2</h2>
  <p>Price: $29.99</p>
</div>
```

```
<div class="product">
  <h2>Product 3</h2>
  <p>Price: $39.99</p>
</div>
```

```
<h1>Recent Articles</h1>
<div class="article">
  <h2>Article 1</h2>
  <p>Lorem ipsum dolor sit amet, consectetur adipiscing elit.</p>
</div>
```

```
<div class="article">
  <h2>Article 2</h2>
  <p>Sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.</p>
</div>
</body>
</html>
```

This HTML snippet represents a simple web page with a list of products and articles. We can use regular expressions to extract information such as product prices and article titles from this HTML content.

Note:

1. Dummy data can be generated as needed for the questions.
2. Chatbot assistance (ChatGPT) is available for support, but final solutions should be provided independently.

Submission Date: 24-12-24

Assignment 5

Note:

- Assignment Submitted in **PDF form**.
- In PDF form, you must include the **screenshot of the code and its output**.
- If not mentioned, then write a Python script that includes the code for each task.
- Include **comments** in your code to explain the purpose and functionality of each step.
- Send it to this email id:
itechsolution93@gmail.com
- You can get help from **ChatGPT** or any **Chatbot** but at the **end**.
- If Any, then It is highly recommended that you read **research papers for assignment**.

Helping Websites for research papers are:

<https://scholar.google.com/>

<https://sci-hub.hkvisa.net/>