|  |  |
| --- | --- |
| **SESSION** | **JUL - AUG 2024** |
| **PROGRAM** | **MASTER OF BUSINESS ADMINISTRATION (MBA)** |
| **SEMESTER** | **IV** |
| **COURSE CODE & NAME** | **DADS404 DATA SCRAPPING** |
|  |  |
|  |  |

**Assignment Set – 1**

**1. Write short notes on different sources of data available for data scraping and factors to choose the right data.**

**Ans 1.**

**Different Sources of Data Available for Data Scraping and Factors to Choose the Right Data**

Data scraping is a method for extracting structured information from various online and offline sources. Identifying the right data sources is crucial for ensuring the quality and relevance of the extracted data. Below are the primary sources of data available for scraping:

**a) Websites** Websites are the most common and accessible sources of data scraping. Publicly available web pages, e-commerce platforms, blogs, and news websites often provide valuable structured or semi-structured data. Websites with APIs make it even easier to scrape data in a

Its Half solved only

Buy Complete assignment from us

**Price – 190/ assignment**

**MUJ Manipal University Complete SolvedAssignments session JULY-AUG 2024**

buy cheap assignment help online from us easily

we are here to help you with the best and cheap help

**Contact No – 8791514139 (WhatsApp)**

**OR**

**Mail us-** [**bestassignment247@gmail.com**](mailto:bestassignment247@gmail.com)

**Our website -** [**www.assignmentsupport.in**](http://www.assignmentsupport.in)

**2. What are the challenges in scraping data manually? Which R packages could help in manual scraping the data?**

**Ans 2.**

**Challenges in Scraping Data Manually and R Packages for Manual Data Scraping**

Scraping data manually involves extracting information without the use of automated tools. While it may be necessary in certain cases, it comes with significant challenges that can impact efficiency and accuracy.

**Challenges in Scraping Data Manually**

**a) Time-Consuming Process** Manual data scraping is highly labor-intensive and time-

**3. Write the steps to scrap the data from any job portal. How python libraries can help in this?**

**Ans 3.**

**Steps to Scrape Data from a Job Portal and the Role of Python Libraries**

Data scraping from job portals involves systematically extracting information such as job titles, company names, locations, and other relevant details. Python, with its versatile libraries, simplifies the entire process. Below are the steps to scrape data from any job portal and an explanation of how Python libraries facilitate this process.

**Steps to Scrape Data from a Job Portal**

**a) Define Objectives and Target Website** The goals and job portal to scrape data should be

**Assignment Set – 2**

**4. Write short notes on API based scrapers. Write benefits and drawback of API based scrapers.**

**Ans 4.**

**API-Based Scrapers: Benefits and Drawbacks**

APIs (Application Programming Interfaces) provide a structured way to access and interact with data hosted on web servers. API-based scrapers utilize these interfaces to retrieve data directly, bypassing the need to scrape web pages manually. This method is widely used for extracting information from platforms like LinkedIn, Twitter, and GitHub.

**Understanding API-Based Scrapers**

API-based scrapers operate by sending HTTP requests to an API endpoint and receiving

**5. What do you understand by data wrangling? What steps or actions come into data wrangling in the industry?**

**Ans 5.**

**Understanding Data Wrangling and Its Steps in the Industry**

**What is Data Wrangling?**

Data wrangling, also known as data munging, is the process of cleaning, organizing, and transforming raw data into a usable format for analysis. It is a critical step in the data pipeline, as raw data collected from various sources often contains inconsistencies, errors, and irrelevant information. Data wrangling ensures that the data is accurate, consistent, and structured, enabling meaningful analysis and informed decision-making.

In the industry, data wrangling involves a combination of technical skills, domain knowledge,

**6. What is the importance of data quality in making decisions? What measures can be taken to improve the quality of data?**

**Ans 6.**

**Importance of Data Quality in Decision-Making**

Any industry needs high-quality data for decision-making. Using erroneous, incomplete, or inconsistent data can cause errors, inefficiencies, and financial losses. For educated judgments, trend prediction, and business process optimization, high-quality data is essential.

Healthcare diagnosis and treatment require precise patient data. In finance, reliable data ensures accurate risk evaluations for investment decisions. In marketing, high-quality data helps