|  |  |
| --- | --- |
| **SESSION** | **FEB MARCH 2025** |
| **PROGRAM** | **MASTER OF BUSINESS ADMINISTRATION (MBA)** |
| **SEMESTER** | **IV** |
| **COURSE CODE & NAME** | **DADS404 DATA SCRAPING** |
|  |  |
|  |  |

**Assignment Set – 1**

**Q1. Write steps to scrap the data from any job portal. How can Python libraries help in this? 6+4**

**Ans 1.**

**Steps to Scrap the Data from Any Job Portal and Use of Python Libraries**

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

Data scraping from a job portal involves a series of organized steps to extract useful information such as job titles, companies, locations, salaries, and job descriptions. The first step is identifying the URL structure of the job portal. This includes analyzing how job listings are presented, paginated, and filtered. Once the target URLs are identified, the next step involves sending HTTP requests using tools like Python’s requests library to fetch the

Its Half solved only

Buy Complete assignment from us

**Price – 190/ assignment**

**MUJ Manipal University Complete SolvedAssignments MARCH 2025**

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

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

**Q2. What are the challenges in scraping data manually? Which R packages could help manually scrap the data? 5+5**

**Ans 2.**

**Challenges in Manual Data Scraping**

Scraping data manually, without automation tools, presents a range of challenges that hinder efficiency, accuracy, and scalability. The first major challenge is time consumption. Extracting data by copying and pasting it from websites is labor-intensive, especially when dealing with large volumes of information across multiple pages or websites. This method is highly inefficient for repetitive tasks.

Another issue is human error. Manual processes are prone to mistakes such as skipping rows,

**Q3. Write short notes on different data sources available for data scraping and factors to choose the right data. 6+4**

**Ans 3.**

**Different Data Sources Available for Data Scraping**

Data scraping can be performed from a variety of sources, each offering unique value depending on the objective. The most common source is websites, especially those with structured content like job listings, product catalogs, and news articles. These websites are typically scraped using HTML parsers and automated scripts. Social media platforms such as Twitter and Facebook are also rich sources of unstructured data. They provide user-generated

**Assignment Set – 2**

**Q4. What is the importance of data quality in making decisions? What measures can be taken to improve the quality of data? 3+7**

**Ans 4.**

**Importance of Data Quality in Decision-Making and Measures to Improve It**

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

High-quality data is the backbone of sound business decisions. When organizations rely on data to shape strategies, set budgets, target customers, or evaluate performance, the accuracy and reliability of that data directly impact the outcome. Poor-quality data—such as duplicated records, missing values, or outdated information—can lead to flawed insights and wrong decisions. For instance, a marketing campaign based on inaccurate demographic data might

**Q5. Write a short note on API based scrapers. Write the benefits and drawbacks of API based scrapers. 5+5**

**Ans 5.**

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

**API-Based Scrapers**

API-based scrapers represent a structured and efficient way to collect data from online platforms. An API, or Application Programming Interface, allows two systems to communicate with each other by sending and receiving requests and responses. Many websites, including social media platforms, job portals, and e-commerce sites, provide public

**Q6. What do you understand by data wrangling? What steps or actions come into data wrangling in the industry? 5+5**

**Ans 6.**

**Data Wrangling and Industry Practices**

**Data Wrangling**

Data wrangling, also known as data munging, refers to the process of transforming raw data into a structured and usable format for analysis. This is a crucial step in data analysis, as raw data is often messy, inconsistent, incomplete, and not readily analyzable. The goal of data wrangling is to make data clean, well-organized, and suitable for downstream tasks such as