

Assignment Booklet

MSCRWEE Programme

M.Sc (Renewable Energy and Environment)

Third Semester (Compulsory)	
MRW-005	Solar Energy and Applications
MRW-006	Bioenergy Conversion and Utilization
MRW-007	Energy Economics and Planning

Third Semester (Electives)	
MRWE-001	Nano Technology in Energy & Environment
MRWE-002	Energy Storage
MEV-021	Introduction to Climate Change
MEVE-001	Environmental Impact Assessment for Environmental Health
MCS-224	Artificial Intelligence and Machine Learning
MCS-226	Data Science and Big Data
MCS-227	Cloud Computing and IoT
MCS-231	Mobile Computing



SCHOOL OF ENGINEERING & TECHNOLOGY
INDIRA GANDHI NATIONAL OPEN UNIVERSITY

Maidan Garhi, New Delhi – 110 068

JANUARY 2025

Dear Student,

Please read the information on assignments in the Programme Guide that we have sent you after your enrolment. A weightage of 30%, as you are aware, has been earmarked for continuous evaluation, **which would consist of one tutor-marked assignment** for this Programme. The assignment for MSCRWEE (Third semester) has been given in this booklet.

Instructions for Formatting Your Assignments

Before attempting the assignment, please read the following instructions carefully:

1) On top of the first page of your answer sheet, please write the details exactly in the following format:

ENROLLMENT NO :

NAME :

ADDRESS :

.....

.....

PROGRAMME CODE:

COURSE CODE:

COURSE TITLE:

STUDY CENTRE:

DATE:

**PLEASE FOLLOW THE ABOVE FORMAT STRICTLY TO FACILITATE
EVALUATION AND TO AVOID DELAY.**

- 2) Use only foolscap size writing paper (but not of very thin variety) for writing your answers.
- 3) Leave 4 cm margin on the left, top and bottom of your answer sheet.
- 4) Your answers should be precise.
- 5) **These assignments submitted should be hand written in your own hand writing.**

We strongly suggest that you should retain a copy of your answer sheets.

- 6) **You cannot fill the Exam Form without** submission of the assignments. So solve it and **submit it at the earliest**. If you wish to appear in the **TEE, June 2025**, you should submit your TMAs by **April 30, 2025**. Similarly, if you wish to appear in the **TEE, December 2025**, you should submit your TMAs by **September 30, 2025**.
- 7) Assignments will be submitted at **your respective regional centre**.

We wish you good luck!

Assignment -6
(To be done **after** studying the course material)

Course Code: MCS-226
Course Title: Data Science and Big Data
Assignment Code: MCS-226/TMA/2025
Maximum Marks: 100

Last Date of Submission: April 30, 2025 (For June TEE), September 30, 2025 (For December TEE)
Note:

- 1. For any question worth 10 marks the word limit is 350 words, for a 20 mark question it is 550 words.**
 - 2. Attempt all questions. All questions carry equal marks.**
-

- Q.1 What is Exploratory Data Analysis (EDA) and why is it important in the data science workflow? What are the key components of the data science process? 10
- Q.2 Discuss the implications of hypothesis testing results in decision-making. Provide examples of real world situations where statistical hypothesis testing is commonly used. 10
- Q.3 Discuss the significance of the three Vs (Volume, Velocity, Variety) in the context of big data. Provide examples of each of the three Vs in real-world scenarios. How does Map Reduce facilitate parallel processing of large datasets? Explain the functionality of the Map function in the Map Reduce paradigm with the help of an example. 10
- Q.4 What is a Data Stream Bloom Filter? Explain its primary purpose in data stream processing. Also, introduce the Flajolet-Martin Algorithm and its role in estimating the cardinality of a data stream. 10
- Q.5 Explain the concept of decision trees in classification. Provide an example of building and visualizing a decision tree using R. How can K-means clustering be applied to a dataset in R? 10
- Q.6 What is data preprocessing, and why is it a crucial step in the data science workflow? Why is it important to identify and handle outliers in a dataset during data preprocessing? 10
- Q.7 Discuss the need for Statistical Hypothesis Testing with the help of an example. Explain types of Errors in Hypothesis Testing. 10
- Q.8 Discuss the Classification, Clustering and Association Rules with different examples. Explain, where we can use Random Forest Algorithm? Use R programming language to discuss Random Forest Algorithm. 10

- Q.9 What is NoSQL database? Discuss how does a Column Database and Document database Work? List and briefly discuss Graph database examples. 10
- Q.10 Explain the process and issues of the following: Advertising on web, Recommendation system, Mining of social networks. 10