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 2026

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 wouldconsist 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 ANDTO 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 2026,** you should submit your TMAs by **April 30, 2026.** Similarly, if you wish to appear in the **TEE, December 2026,** you should submit your TMAs by **September 30, 2026.**

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/2026
Maximum Marks: 100

Last Date of Submission: April 30, 2026 (For June TEE), September 30, 2026 (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	Define the term data science. Describe its applications in two industries of your choice (e.g., healthcare, finance, e-commerce). What role does the data science lifecycle play in managing data projects?	10
Q.2	Explain Exploratory Data Analysis (EDA) and its importance. What are the main steps in performing EDA on a new dataset? Describe two methods for detecting outliers and how handling outliers impacts data analysis.	10
Q.3	Describe the role of statistical hypothesis testing in data analysis. What are Type I and Type II errors, and how do they affect decision-making? Provide an example of hypothesis testing in a real-world scenario.	10
Q.4	Discuss the 4 Vs of big data (Volume, Velocity, Variety, and Veracity). Provide a real-world example of each, explaining how these characteristics create challenges in big data management.	10
Q.5	Explain the Hadoop architecture with a focus on HDFS and the master/slave architecture. How do NameNode and DataNodes work together to store and manage large datasets? Provide a basic example of this storage process.	10
Q.6	Compare Apache Spark, Hive, and HBase in terms of functionality, data processing methods, and use cases. When would Spark be preferred over traditional MapReduce, and why?	10
Q.7	Describe the purpose and functionality of a *Bloom filter* in data stream processing. How does the Bloom filter efficiently check for element presence? Describe the Flajolet-Martin algorithm for cardinality estimation in data streams.	10
Q.8	What is the PageRank algorithm, and how is it used in link analysis? Describe the concept of “flow of rank” in PageRank. Explain how the PageRank of a web page is calculated using the flow model.	10

Q.9	Discuss the challenges of online advertising and recommendation systems. Explain the concept of collaborative filtering with an example, and discuss the role of clustering in social network analysis.	10
Q.10	What is the Random Forest algorithm? Explain how it can be applied to classification problems. Write a program in R to implement a Random Forest classifier on a sample dataset and explain its output.	10