Teaching Plan

DSE 3

Introduction to Data Science

Instructor's Name: Rabisankar Pramanik, State Aided College Teacher, Debra Thana SKS Mahavidyalaya.

Course Code: DSE 3	**Type of Course:** Discipline Specific Electives
Credits: 06 (Theory-04 and Practical-02)	**Duration:** February 06, 2023 to June 14, 2023
Semester: Six	**Marks:** 75 [**End Semester Exam:** 60 (**Theory:** 40 and **Practical:** 20) + **Internal Assessment:** 10 + **Class Attendance:** 5]

Intended Audience: B Sc Computer Science undergraduate students.

Pre-Requisites: Basic knowledge of mathematics, statistics and databases. Familiarity with any programming language.

Course Abstract:

This is an introductory course on Data Science. The objective of the course is to provide an exposition first to the data science, then to the tools like Git, GitHub, R and RStudio. Gain hands-on familiarity with R programming language. Getting data. Data analysis and data visualization. Concepts and tools for modern data analysis and R markdown.

Skill you will get:

- **Data Science**
- **R programming**
- **Git, GitHub**
- **RStudio**

Course Plan:

Theory:

Unit	Topics	No of lectures	Duration in Hours
Unit 1	**Data Scientist's Tool Box**: Turning data into actionable knowledge, introduction to the tools that will be used in building data analysis software: version control, markdown, git, GitHub, R, and RStudio.	4	4
Unit 2	**R Programming Basics**: Overview of R, R data types and objects, reading and writing data, Control structures, functions, scoping rules, dates and times, Loop functions, debugging tools, Simulation, code profiling.	10	10
Unit 3	**Getting and Cleaning Data**: Obtaining data from the web, from APIs, from databases and from colleagues in various formats. basics of data cleaning and making data ―tidy.	6	6
Unit 4	**Exploratory Data Analysis**: Essential exploratory techniques for summarizing data, applied before formal modeling commences, eliminating or sharpening potential hypotheses about the world that can be addressed by the data, common multivariate statistical techniques used to visualize high dimensional data.	14	14
Unit 5	**Reproducible Research**: Concepts and tools behind reporting modern data analyses in a reproducible manner, To write a document using R markdown, integrate live R code into a literate statistical program, compile R markdown documents using knitr and related tools, and organize a data analysis so that it is reproducible and accessible to others.	16	16

Practical:

Assignment	Topics	No of lectures	Duration in Hours
Assignment 1	Write a program that prints “Hello World” to the screen.	1	2
Assignment 2	Write a program that asks the user for a number n and prints the sum of the numbers 1 to n	1	2
Assignment 3	Write a program that prints a multiplication table for numbers up to 12.	1	2
Assignment 4	Write a function that returns the largest element in a list.	1	2
Assignment 5	Write a function that computes the running total of a list.	1	2
Assignment 6	Write a function that tests whether a string is a palindrome.	1	2
Assignment 7	Implementation the following sorting algorithms: Selection sort, Insertion sort, Bubble Sort	3	6
Assignment 8	Implementation linear search.	1	2
Assignment 9	Implementation binary search.	1	2
Assignment 10	Implementation matrices addition, subtraction and Multiplication.	1	2
Assignment 11	Finding mean, median and mode from given data.	2	4
Assignment 12	Implementing basic plotting.	2	4

Textbooks:

1. **Doing Data Science: Straight Talk from the Frontiline** by **Rachel Schutt, Cathy O'Neil, O'Reilly**, 2013.

2. **Data Science for Business – What You Need to Know About Data Mining and Data-Analytic Thinking** by **Foster Provost, Tom Fawcett**, O'Reilly, 2013.

* * *