

Master of Engineering (M.E)

Semester : I

Branch: Computer Engineering (Software Engineering)

Course Code:	20-CE-PG-049010104
Course Name:	Fundamentals of Data Science
Category of Course:	Core
Prerequisite Course:	Data Structures, Basics of Probability and Statistics, Data Mining

Teaching Scheme				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Total Hours
3	0	2	4	40

Course Objectives	
1	To understand the underlying fundamental concepts and techniques of Data Science.
2	Develop a broad academic and practical literacy in computer engineering, statistics, and optimization, with relevance in data science, so that students are able to critically select and apply appropriate methods and techniques to extract relevant and important information from data.
3	Enable students to understand not only how to apply certain methods, but when and why they are appropriate.
4	Expose students to real-world problems in the classroom and through experiential learning.
5	Provide strong core training so that graduates can adapt easily to changes and new demands from industry.
6	To gain experience of doing independent projects study and research.

Syllabus			
Unit No.	Topic	Prerequisite Topic	Teaching Hours
01	Introduction of Data Science	---	4 (8%)
	1.1 Why Data Science?, Difference Between Business Intelligence and Data Science		
	1.2 Lifecycle or Process model of a Data Science		
	1.3 Application and Challenges of Data science Technology		
02	Collect and Manipulate Data from Different Data Sources	---	4 (10%)
	2.1 Sources of data, Data Collection and APIs		
	2.2 Exploring and Fixing Data		
	2.3 Data Storage and Management		
03	Data Analysis	---	5 (10%)
	3.1 Fundamentals of Descriptive Statistics and Probability		
	3.2 Measures of Central Tendency		
04	Machine Learning Algorithms	---	5 (10%)
	4.1 Learning from Data with Your Machine		
	4.2 Principal Component Analysis (PCA) and its application		
	4.3 Linear Discriminant Analysis (LDA)		
05	Creating Data Visualizations	---	4 (10%)
	5.1 Types of Visualizations		
	5.2 Picking the Most Appropriate Design Style		
	5.3 Add Context, Choosing the Most Appropriate Data Graphic Type		
06	D3.js for Data Visualization	---	3 (12%)
	6.1 Introducing the D3.js library		
07	Web-Based Applications for Visualization Design	---	4 (10%)
	7.1 Using Collaborative Data Visualization Platforms		
	7.2 Visualizing Spatial Data with Online Geographic Tools		
	7.3 Visualizing with Open Source		
08	Computing for Data Science	---	3 (10%)
	8.1 Using Python for Data Science, Understanding Basic Concepts in Python		
	8.2 Getting on a First-Name Basis with Some Useful Python Libraries		
09	Using Open Source R for Data Science	---	4 (12%)
	9.1 Introducing the Fundamental Concepts, Previewing R Packages		
	9.2 Using SQL in Data Science, Getting Started with SQL		
10	Recent Trends		4

10.1 Free Data Science Tools and Applications	---	(8%)
10.2 Application Development Methods of used in Data Science		
10.3 Using Data Science to Describe and Predict Criminal Activity		

Course Outcome

1	Understand How data is collected, managed and stored for data science.
2	Understand Data Science Project Lifecycle.
3	Understand the fundamentals of statistics.
4	How to Visualize Data and Perform Exploratory Data Analysis.
5	Understand the mechanics of regression analysis.
6	Get understanding on conditions and loops, functions in R, objects, classes, and debugging and Master R programming and understand how various statements are executed in R.

Suggested Reference Books

1	The Data Science Handbook, Field Cady, Wiley
2	Data Science, John D Kellehar, MIT Press
3	Doing Data Science, Cathy O’Neil and Rachel Schutt, Straight Talk From The Frontline. O’Reilly.
4	Introduction to Data Science, Davy Cielen, Arno D B Meysman and Mohamed Ali, Manning, dreamtech press
5	Practical Data Science, Nina Zumwl and John Mount, Manning, dreamtech press

Proposed Evaluation Scheme by Academicians (Percentage of Weightage out of 100%)

Theory Descriptive Test	<input type="text"/>	MCQ Test	<input type="text"/>	Hands on Project	<input type="text"/>
Formulas and Derivation Test	<input type="text"/>	Numerical Test	<input type="text"/>	Seminar	<input type="text"/>

Practical Project/Hands On Project

Sr. No.	List of Practical Projects	Linked with Unit
1	Explain Lifecycle or Process model of a Data Science. Implement Lifecycle or Process model using suitable Example.	Unit 1
2	Implement Data Science project is looking at the common forms of fake news. These days, it’s hard enough for the average social media user to determine when an article is made up with an intention to deceive. So is it possible to build a model that can discern whether a news piece is credible. Hint: Two common forms of fake news to focus on: clickbait (“shocking headlines meant to generate clicks to increase ad revenue”) and propaganda (“intentionally misleading or deceptive articles meant to promote the author’s agenda”).	Unit 2
3	Compute the Measures of Central Tendency Consider the following data points. 17, 16, 21, 18, 15, 17, 21, 19, 11, 23	Unit 3
4	Implement Regression Problem using R programming. You should take Database from: UCI Machine Learning repository. You should always check missing values in data and check correlation matrix.	Unit 4
5	Implement PCA using Python Programming.	Unit 4
6	Implement problem of Credit Card Fraud Detection mini Project in R.	Unit 5
7	Music Recommendation System Project using Python and R	Unit 6
8	Implement Driver Drowsiness detection in Python.	Unit 8
9	Implement problem of data science project in python to build a predictive model and find out the sales of each product at a given Big Mart store.	Unit 8
10	Demonstrate the SQL Techniques to Perform Data Analysis for Analytics and Data Science.	Unit 9

List of Recommended MOOC Courses:

- 1) <https://www.coursera.org/learn/python-for-data-visualization>
- 2) <https://www.coursera.org/learn/python-data-analysis>
- 3) https://onlinecourses.nptel.ac.in/noc21_cs33/preview
- 4) <https://www.mooc-list.com/course/what-data-science-coursera>