

## **Short Syllabus**

**BCSE206L**

**Foundations of Data Science**

**(3-0-0-3)**

Data Science Context - Need for Data Science, Data Science Process; Databases for Data Science - Data Munging, Filtering, Joins, Aggregation, Window Functions, Ordered Data, No-SQL, Document Databases; Data Science Methodology- Analytics for Data Science, Examples of Data Analytics, Data Analytics Lifecycle; Data Analytics on Text- Major Text Mining Areas, Text analytics tasks; Platform for Data Science - Python Libraries, Exploration Data Analysis; GNU Octave for Mathematical Operations - Handling Vectors and Matrices, Eigen Vectors and Eigen Values; Tableau - Dashboard Design Principles, Special Chart Types.

Course code	Course Title	L	T	P	C
BCSE206L	Foundations of Data Science	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			
Course Objectives					
<div><div></div><div><div>1.</div><div>To provide fundamental knowledge on data science with querying and analytics required for the field of data science.</div></div><div><div>2.</div><div>To understand the process of handling heterogeneous data, pre-process and visualize them for better understanding.</div></div><div><div>3.</div><div>To gain the fundamental knowledge on data science tools and gain basic skill set to solve real-time data science problems.</div></div></div>					
Course Outcome					
<div>Upon completion of the course the student will be able to</div> <div><div></div><div><div>1.</div><div>Ability to obtain fundamental knowledge on data science.</div></div><div><div>2.</div><div>Demonstrate proficiency in data analytics.</div></div><div><div>3.</div><div>Apply advanced tools to work on dimensionality reduction and mathematical operations.</div></div><div><div>4.</div><div>Handle various types of data and visualize them using through programming for knowledge representation.</div></div><div><div>5.</div><div>Demonstrate numerous open source data science tools to solve real-world problems through industrial case studies.</div></div></div>					
Module:1	Data Science Context	5 hours			
Need for Data Science – What is Data Science - Data Science Process – Business Intelligence and Data Science – Prerequisites for a Data Scientist – Tools and Skills required.					
Module:2	Databases for Data Science	7 hours			
Structured Query Language (SQL): Basic Statistics, Data Munging, Filtering, Joins, Aggregation, Window Functions, Ordered Data, preparing No-SQL: Document Databases, Wide-column Databases and Graphical Databases.					
Module:3	Data Science Methodology	8 hours			
Analytics for Data Science – Examples of Data Analytics – Data Analytics Lifecycle: Data Discovery, Data Preparation, Model Planning, Model Building, Communicate Results.					
Module:4	Data Analytics on Text	7 hours			
Major Text Mining Areas – Information Retrieval – Data Mining – Natural Language Processing NLP) – Text analytics tasks: Cleaning and Parsing, Searching, Retrieval, Text Mining, Part-of-Speech Tagging, Stemming, Text Analytics Pipeline. NLP: Major components of NLP, stages of NLP, and NLP applications.					
Module:5	Platform for Data Science	6 hours			
Python for Data Science –Python Libraries – Data Frame Manipulation with numpy and pandas – Exploration Data Analysis – Time Series Dataset – Clustering with Python – Dimensionality Reduction. Python integrated Development Environments (IDE) for Data Science.					
Module:6	GNU Octave for Mathematical Operations	6 hours			
Handling Vectors and Matrices: Multiplication, Transpose, Random Matrix creation, Eigen Vectors and Eigen Values, Determinants. Arithmetic Operations – Set Operations – Plotting Data.					
Module:7	Tableau	4 hours			
Tableau Introduction – Dimensions, Measures, Descriptive Statistics, Basic Charts, Dashboard Design Principles, Special Chart Types, Integrate Tableau with Google Sheets.					
Module:8	Contemporary Issues	2 hours			

	<b>Total Lecture hours:</b>		<b>45 hours</b>	
<b>Text Book(s)</b>				
1.	Sanjeev Wagh, Manisha Bhende, Anuradha Thakare, 'Fundamentals of Data Science, CRC Press, 1 <sup>st</sup> Edition, 2022.			
<b>Reference Books</b>				
1.	Avrim Blum, John Hopcroft, Ravindran Kannan, "Foundations of Data Science", Cambridge University Press, First Edition, 2020.			
2.	Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media, 1 <sup>st</sup> Edition, 2015.			
3.	Ani Adhikari and John DeNero, 'Computational and Inferential Thinking: The Foundations of Data Science', GitBook, 2019.			
Mode of Evaluation : Continuous Assessment Tests, Quizzes, Assignment, Final Assessment Test				
Recommended by Board of Studies		12-05-2022		
Approved by Academic Council		No. 66	Date	16-06-2022