



Syllabus

Programme Name: B. Tech.		Session:
Course Code: CSD101	Course Name: Introduction to Data Science	Semester: I

Credits (Total)	L	T	P	Marks (Internal/External)		Contact Hours (per week)	Independent Study Hour (per week)	Section (Group)
3	3	0	0	30	70	3	3	Pls name the section if an-y
Curriculum level						Basic and applied	Student specific course outcome	Higher Education Placement Research

Course Objective:

The objective of this course is to provide a basic understanding of data science field and its implementation in Various Industries.

Course outcomes: After completion of course, the student will be able to:

CO-1	Understand the evolution and significance of Data Science in the business world.
CO-2	Evaluate challenges and scope in Data Science project management, comparing project methods for improved success.
CO-3	Apply statistical measures and techniques in Data Science, including linear regression and probability concepts.
CO-4	Recognize the role of computer science, databases, and data warehousing in Data Science.
CO-5	Explore real-world applications of Data Science and understand industry use cases

Teaching Pedagogy:

T1	Classroom teaching (white board), Power Point Presentations, Interactive lectures, Inquiry based teaching
T2	ABL activities, Assignments, Flip Class/ Seminars, Quiz, Oral Viva-voce examination

Assessment Tools

AT1-1	Quiz
AT1-2	Activity Based Learning
AT1-3	Midterm Exams
AT1-4	Flip Class

AT1-5	Seminar Presentation
AT1-6	Assignments
AT1-7	Poster
AT1-8	Oral Viva-voce examination
AT1-9	Industrial Visit Report

Prerequisites: Basic understanding of Data.

Module wise contents details	Assessment tools
Module I: Introduction : (5 Hours) Introduction to Data Science, Definition and description of Data Science, history and development of Data Science, terminologies related with Data Science, basic framework and architecture, difference between Data Science and business analytics, importance of Data Science in today's business world, primary components of Data Science, users of Data Science and its hierarchy.	Quiz Mid-term Exam Assignment
Module II: Data Science Project Management (5 Hours) Data Science project framework, Stages in a Data Science Project ,execution flow of a Data Science project, various components of Data Science projects, stakeholders of Data Science project, , challenges and scope of Data Science project management, process evaluation model, comparison of Data Science project methods, improvement in success of Data Science project models.	Mid-Term Quiz Assignment
Module III: Mathematics behind Data Science: (9 Hours) Role of mathematics in Data Science, importance of probability and statistics in Data Science, important types of statistical measures in Data Science : Descriptive, Predictive and prescriptive statistics, introduction to statistical inference and its usage in Data Science, application of statistical techniques in Data Science, Basics of probability, permutation and combination, introduction to linear Regression model, mean, mode, median, Outliers, Leverage points, Business Logics, Feature Engineering, bad data identification and correction.	Mid-Term Oral Viva-voce examination Seminar Presentation
Module IV: Computers in Data Science(6 Hours) Role of computer science in Data Science, various components of computer science being used for Data Science, role of relation data base systems in Data Science: SQL, NoSQL, role of data warehousing in Data Science, terms related with data warehousing techniques, importance of operating concepts and memory management, various freely available software tools used in Data Science : R, Python, important proprietary software tools, different business intelligence tools and its crucial role in Data Science project presentation.	Quiz Assignment Industrial Visit Report Seminar Presentation
Module V: Applications of Data Science: (5 Hours) Applications of Data Science in various fields. industry use cases of Data Science implementation General use cases of data science in Finance-	Quiz Assignment

defaulter detection, E-Commerce-Recommendation Systems, Banking Industry-Loan credibility System, Real Estate, and GIS Systems- optimal route founding (Olla, Uber)	Industrial Visit Report Poster Oral Viva-voce examination
---	--

Additional Learning:

List of Assignments	<ul style="list-style-type: none"> • Define Data Science and provide a historical overview of its development. • Explain the key terminologies associated with Data Science. • Highlight the differences between Data Science and business analytics. • Discuss the primary components of Data Science and their importance in today's business world. • Describe the hierarchy of users involved in Data Science. • Outline the stages involved in a Data Science project and explain the execution flow. • Identify and analyze the various components of Data Science projects. • Discuss the challenges faced in Data Science project management and propose solutions. • Compare different Data Science project methods and their impact on project success. • Explore the role and importance of stakeholders in a Data Science project. • Elaborate on the role of mathematics in Data Science and its significance. • Differentiate between descriptive, predictive, and prescriptive statistics in Data Science. • Explain the application of statistical techniques in Data Science with relevant examples. • Discuss the basics of probability, permutation, and combination in the context of Data Science. • Explore the importance of linear regression, mean, mode, median, outliers, and leverage points in Data Science. • Investigate the role of computer science in Data Science and its various components. • Compare and contrast SQL and NoSQL in the context of Data Science. • Discuss the significance of data warehousing techniques and their terms. • Evaluate the importance of operating concepts and memory management in Data Science. • Analyze the role of software tools like R, Python, and proprietary tools in Data Science projects. • Provide examples of Data Science applications in the finance industry, focusing on defaulter detection. • Explain how recommendation systems are implemented in e-commerce using Data Science. • Discuss the role of Data Science in determining loan credibility in the banking industry. • Explore Data Science applications in real estate, highlighting its impact. • Describe how GIS systems utilize Data Science for optimal route finding in services like Olla and Uber.
Suggested reading:	<p>Texts:</p> <ul style="list-style-type: none"> • Think Python by Allen B Downey • Cathy O’Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O’Reilly. 2014. • Avrim Blum, John Hopcroft and Ravindran Kannan. Foundations of Data Science. <p>References:</p> <ul style="list-style-type: none"> • Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press. 2014. (free online) • Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. ISBN

	<p>0262018020. 2013.</p> <ul style="list-style-type: none"> • Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking. ISBN 1449361323. 2013. • Trevor Hastie, Robert Tibshirani and Jerome Friedman. Elements of Statistical Learning, Second Edition. ISBN 0387952845. 2009. (free online)
Suggested e- resources (Websites/e-books)	<p>https://www.coursera.org/specializations/introduction-data-science</p> <p>https://pg-p.ctme.caltech.edu/data-science-course-certification</p> <p>https://cloudxlab.com/course/188/pg-certificate-program-in-data-science-ai-by-cec-iit-roorkee</p> <p>Website:</p> <p>https://www.heavy.ai/learn/data-science</p>

Assessment Plan:

Component of Evaluation	Description	Code	Weightage %
Continuous Internal Evaluation	Mid Term	CT	15%
	Seminar/Viva-Voce/Quiz/Home Assignment	S/V/Q/HA	10%
Attendance	A minimum of 75% Attendance is required to be maintained by a student to be qualified for taking the End Semester examination. The dispensation of 25% includes all types of leaves. including medical leaves.	A	5%
End Semester Examination	End Semester Examination	ESE	70%
Total			100%

Abbreviations: CT: Class Test, HA: Home Assignment, S/V/Q: Seminar/Viva/Quiz, ESE: End Semester Examination; A: Attendance

Course Articulation Matrix (Mapping of COs with POs)

Course Outcomes	Correlation with POs												Correlation with PSOs		
	P O 1	P O 2	P O 3	P O 4	P O 5	P O 6	P O 7	P O 8	P O 9	P O 10	P O 11	P O 12	P S O 1	P S O 2	P S O 3
CO1	3	3	1	3	1				2		2	1			
CO2	3	2	2	2	2				2		1	1			
CO3	3	2	2	2	2				3		3	1			
CO4	3	3	2	3	2				1		2	1			
CO5	2	2	1	2	3				2		2	1			