



# K. S. INSTITUTE OF TECHNOLOGY

An Autonomous Institution under VTU, Approved by AICTE  
Department of Computer Science & Engineering  
**M.Tech FIRST SEMESTER SYLLABUS**

<b>Course: DATA SCIENCE AND MANAGEMENT</b>		Semester	I
<b>Course Code</b>	<b>25MCS102</b>	CIE Marks	50
Teaching Hours/Week (L:P:SDA)	4:0:2	SEE Marks	50
Total Hours of Pedagogy	50	Total Marks	100
Credits	03	Exam Hours	03
Examination type (SEE)	<b>Theory</b>		

## Course Objectives (Course Skill Set)

1. Explain the foundational concepts of data science, including its history, significance, and the data science process.
2. Apply statistical methods and data analysis techniques to interpret and draw insights from complex datasets.
3. Implement various machine learning algorithms and assess their performance using appropriate evaluation metrics in real-world scenarios.
4. Utilize data visualization tools and techniques to effectively communicate findings and insights to diverse audiences.

### Module-1

**Module 1:** Introduction to Data Science and R Tool, Overview of Data Science Importance of Data Science in Engineering , Data Science Process , Data Types and Structures, Introduction to R Programming, Basic Data Manipulation in R, Simple programs using R. Introduction to RDBMS: Definition and Purpose of RDBMS Key Concepts: Tables, Rows, Columns, and Relationships, SQL Basics: SELECT, INSERT, UPDATE, DELETE Importance of RDBMS in Data Management for Data Science

### Module-2

**Module 2:** Linear Algebra for Data Science, Algebraic View, Vectors and Matrices, Product of Matrix & Vector, Rank and Null Space, Solutions of Over determined Equations, Pseudo inverse, Geometric View, Vectors and Distances, Projections, Eigenvalue Decomposition.

### Module-3

**Module 3:** Statistical Foundations, Descriptive Statistics, Notion of Probability, Probability Distributions Understanding Univariate and Multivariate Normal Distributions, Mean, Variance, Covariance, and Covariance Matrix, Introduction to Hypothesis Testing, Confidence Intervals for Estimates.

### Module-4

**Module 4:** Optimization and Data Science Problem Solving, Introduction to Optimization Understanding Optimization Techniques, Typology of Data Science Problems, Solution Framework for Data Science Problems.

### Module-5

**Module 5:** Regression and Classification Techniques, Linear Regression, Simple Linear Regression and Assumptions, Multivariate Linear Regression, Model Assessment and Variable Importance, Subset Selection, Classification Techniques, Classification using Logistic Regression.

**Course outcome (Course Skill Set)**

At the end of the course the student will be able to:

- CO 1. Identify and demonstrate data using visualization tools.
- CO 2. Make use of Statistical hypothesis tests to choose the properties of data, curate and manipulate data.
- CO 3. Utilize the skills of machine learning algorithms and techniques and develop models.
- CO 4. Demonstrate the construction of decision tree and data partition using clustering.
- CO 5. Experiment with social network analysis and make use of natural language processing skills to develop data driven applications.

**Assessment Details (both CIE and SEE)**

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 40% of the maximum marks (20 marks). A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 35% (18 Marks out of 50) in the semester-end examination (SEE), and a minimum of 40% (40 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together

**Continuous Internal Evaluation:**

1. Three Unit Tests each of **25 Marks**
2. Two assignments each of **25 Marks** or one Skill Development Activity of **25 marks** to attain the COs and POs

The sum of three tests, two assignments/skill Development Activities, will be scaled down to **50 marks**.

**CIE methods /question paper is designed to attain the different levels of Bloom's taxonomy as per the outcome defined for the course.**

**Semester End Examination:**

1. The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.
2. The question paper consists of Part A and Part B. Part A consists of 10 questions from 5 modules, each carrying 2 marks.
3. Part B consists of 10 questions. Each full question is for 16 marks. There will be two full questions (with a maximum of three sub-questions) from each module.
4. Each full question will have a sub-question covering all the topics under a module.
5. The students will have to answer five full questions, selecting one full question from each module.

**Suggested Learning Resources:****Text Books:**

1. "Python for Data Analysis" by Wes McKinney, 2nd Edition (2018)
2. "Data Science from Scratch: First Principles with Python" by Joel Grus, 2nd Edition (2019)

**Reference Books**

1. "An Introduction to Statistical Learning" by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, 2nd Edition (2021)

2. "The Elements of Statistical Learning" by Trevor Hastie, Robert Tshigami, and Jerome Friedman, 2<sup>nd</sup> Edition (2009).

3. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett, 2nd Edition (2013)

**Web links and Video Lectures (e-Resources):**

1. <https://www.coursera.org/specializations/jhu-data-science>
2. <https://www.kaggle.com/learn/data-science>
3. <https://www.edx.org/professional-certificate/harvardx-data-science>
4. <https://www.youtube.com/playlist?list=PL4cUxeGkcC9g1s4L6G8p8Fq5XK6Pq7b1k>

**Skill Development Activities Suggested:** The students with the help of the course teacher can take up relevant technical activities which will enhance their skill. The prepared report shall be evaluated for CIE marks.

**Course outcome (Course Skill Set)**

At the end of the course the student will be able to:

Sl. No.	Description	Blooms Level
CO1	Explore the foundational concepts of data science, history, significance, and process.	L3
CO2	Apply statistical methods and data analysis techniques to interpret and draw insights from complex datasets.	L3
CO3	Implement various machine learning algorithms and assess their performance using appropriate evaluation metrics in real-world scenarios.	L2
CO4	Utilize data visualization tools and techniques to effectively communicate findings and insights to diverse audiences.	L4