SEMESTER	3
Paper Number	MCMS 4315
Paper Title	Foundations of Data Science
No. of credits	6
Theory / Composite	Composite
No. of periods assigned	Th: 4 Pr: 4
Name of faculty member(s)	
Course description / objectives	On completion of this course, the students will be able to: 1. understand the various roles of Data Science and Data Scientists along with various applications. 2. learn how to deal with Big Data. 3. identify the various facets of Exploratory Data Analysis and implement them using suitable programming language. 4. learn the mathematics relevant to data science like linear algebra, probability and optimization techniques and their role in advanced courses of data science. 5. introduce to the basic learning techniques and their applications.
Syllabus	**Theory - 60 marks** Introduction to Data Science: Data Science Lifecycle; Pre–requisites of Data Science; Role of Data Science and Data Scientist; Stages in a Data Science Project; Applications of Data Science; Data Security Issues. Dealing with Big Data: Big Data and Cloud technologies ; Introduction to HADOOP; Big Data, Apache Hadoop, MapReduce; Data Serialization; Data Extraction; Stacking Data; Dealing with data. Exploratory Data Analysis (EDA): Statistical Measures, Basic tools of EDA, Discovery. Basic Functionalities of a data object; Merging of Data objects; Concatenation of data objects; Types of Joins on data objects; Exploring a Dataset; Analysing a dataset; Group By operations; Aggregation; Concatenation; Merging ; Joining. Statistics and Probability Distributions: Data collection and classification; Graphical representation; Measures of central tendency and dispersion; Moments, Skewness, Kurtosis; Theory of Probability; Random Variable: Discrete and Continuous Probability Distributions, Joint Distributions; Theoretical Distributions: Binomial, Poisson and Normal; Sampling Theory; Hypothesis Testing. Correlation and Regression: Curve fitting, Method of Least Squares and their application for various types of curves; Scatter Diagram; Correlation; Standard Error; Linear regression; Multiple Regression; Rank Correlation. Basics of Linear Algebra: Vectors: Linear dependence and independence; Eigen Values and Eigen Vectors: properties; Caley-Hamilton theorem and its application; Reduction to Diagonal form; Quadratic form and their nature; Principal axes form. Learning techniques: Supervised learning; Unsupervised learning; Modelling Process; Training, Validating and Testing; Overview of Deep Learning. **Lab - 40 marks**
Reading/Reference Lists	1. Jojo Moolayil, “Smarter Decisions : The Intersection of IoT and Data Science”, PACKT, 2016. 2. Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly, 2015. 3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013.

	4. Raj, Pethuru, "Handbook of Research on Cloud Infrastructures for Big Data Analytics", IGI Global. 5. Ethem Alpaydin, "Introduction to Machine Learning". MIT Press. 2020.
Evaluation	Total – 100 (Theory – 60, Practical – 40) Theory – CIA – 10 Semester Examination – 50