

Semester	6
Course	MAJOR
Paper Code	
Paper Title	FOUNDATIONS OF DATA SCIENCE
No. of Credits	4
Theory / Practical / Composite	COMPOSITE
Minimum No. of preparatory hours per week a student has to devote	5
Number of Modules	ONE
Syllabus	Introduction to Data Science: Introduction to data – structured, unstructured, semi-structured, data sets & patterns, Introduction to Data Science, ML Development Life Cycle, Importance of Data Science, Differences between AI, ML, DL, Data Science & Data Analytics, Real world applications of data science, Steps in data science process, case study based on real life applications, Exploratory Data analysis – build the model. Data Preprocessing: An Overview - Data Quality, Need to preprocess the data. Major Tasks in Data Preprocessing. Data cleaning - Missing Values Noisy Data, Data Cleaning as a Process, Data Integration, Data Reduction, Data transformation and Data Discretization. Data Visualization and EDA. Basic Statistics: Data collection and classification; Graphical representation; Measures of central tendency and dispersion; Moments, Skewness, Kurtosis. Probability: Basic concepts, Conditional probability, Independence, Bernoulli Trials and Binomial Distribution, Random Variables, Bayes Theorem, Expectation, Variance, Discrete and Continuous Probability Distributions, Uniform, Poisson and Normal Distributions, Z-score. Correlation and Regression: Curve fitting, Method of Least Squares and their application for various types of curves, Goodness of Fit, Correlation, Scatter Diagram, Linear regression; Multiple Regression; Rank Correlation. Evaluation: Evaluating model performance-Confusion matrices, Precision and recall, Sensitivity and specificity, F-measure, ROC curves, Cross validation, K-fold cross validation, Bootstrap sampling. Hyperparameter tuning.
Learning Outcomes	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 basic learning techniques and their applications.	
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	Theory CIA: 12 Attendance: 3 Semester Exam: 45	Practical CA: 38 Attendance: 2
Paper Structure	Answer 3 out of 5 of 15 marks each	