Semester: VI	
Programme : B.Sc. Computer Science (Hons)	
Course : FOUNDATIONS OF DATA SCIENCE	
Paper code: C3CS230632T / C3CS230632P	**Credits: 4**
Hours/week : Theory: 3 / Practical 2	
Category: Core/MDC/SEC/VAC : Core	
Theory / Practical / Composite : Composite	
No of Modules : 1	

Course Overview: This course provides a comprehensive introduction to Data Science, focusing on the concepts, methods, and tools required to analyze and extract insights from data. It begins with an overview of data types, datasets, and patterns, emphasizing the importance of data science and its role in data analytics and machine learning applications. The course covers the complete machine learning development life cycle, exploratory data analysis, and model-building fundamentals. Learners gain hands-on exposure to data preprocessing techniques, including data cleaning, integration, feature transformation, normalization, dimensionality reduction, and visualization to ensure data quality and readiness for analysis. The mathematical foundations necessary for data science are introduced through linear algebra, descriptive statistics, and probability theory, including hypothesis testing and probability distributions. Further, the course explores correlation and regression techniques for data modeling and interpretation. Real-life case studies are integrated throughout the course to reinforce concepts and enable students to apply theoretical knowledge to practical data-driven problem-solving scenarios.

Course Outcome:

1.**Recall and explain** - Recall fundamental concepts of data types, data science definitions, machine learning life cycle, and statistical and probability theories. Explain the importance and applications of data science and differentiate between key related fields such as AI, ML, and data analytics.

2.**Apply** data preprocessing techniques such as cleaning, integration, transformation, normalization, and visualization on real-world datasets to prepare data for analysis.

3.**Analyze** datasets using descriptive statistics, exploratory data analysis, probability distributions, correlation, and regression methods to identify patterns and relationships.

4.**Design** and implement predictive models using regression techniques, curve fitting, and feature engineering to solve practical data-driven problems.

5.**Evaluate** the performance of machine learning models and data analysis approaches using metrics such as goodness-of-fit, hypothesis testing, and model validation techniques.

6.**Create** end-to-end data science solutions by integrating data collection, preprocessing, statistical analysis, modeling, evaluation, and visualization for real-life case studies.

Prerequisites: Basic Statistics

SYLLABUS

UNIT/Module	CONTENT	HOURS	CO Mapping	COGNITIVE LEVEL
I.	Introduction to data – structured, unstructured, semi-structured, data sets & amp; patterns, Introduction to Data Science, ML Development Life Cycle, Importance of Data Science, Data Analytics, Real-world applications of data science in Machine Learning Algorithms,	6	CO1,CO2, CO3, CO6	K1, K2, K3, K4, K6 (Remember, Explain, Apply, Analyze, Create)

	Exploratory Data analysis – build the model.			
II.	**Data Pre-processing:** An Overview - Data Quality, Need to pre-process the data. Major Tasks in Data Pre-processing. Data cleaning – Missing Values Noisy Data, Data Integration, Feature Transformation, Construction, Selection and Extraction methods, Data Discretisation, Normalisation and Standardisation, Dimensionality Reduction, Visualisation and EDA.	6	CO2, CO3 CO4, CO6	K3, K4, K6 (Apply, Analyze, Create)
III.	**Basics of Linear Algebra:** Vectors: Linear dependence and independence; Eigen Values and Eigen Vectors: Properties and Applications.	4	CO1, CO4	K1, K2, K6 (Recall, Explain, Design)
IV.	**Descriptive Statistics:** Data collection and classification; Measures of central tendency and dispersion; Moments, Skewness, Kurtosis.	4	CO1, CO3	K1, K4 (Recall, Analyse)
v.	**Probability 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, basics of hypothesis testing	10	CO1, CO3, CO5	K1,K4, K5 (Recall, Analyze, Evaluate)
VI.	**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. Case study based on real-life applications	9	CO4, CO5 CO6	K5, K6 (Design, Create)

Text Books
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.
Suggested readings
1. Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global. 2. Ethem Alpaydin, “Introduction to Machine Learning“. MIT Press. 2020.
Web Resources

1. Data Science for Engineers, IIT Madras, Prof. Ragunathan Rengasamy, Prof Shankar Narasimhan, https://nptel.ac.in/courses/106106179		
Evaluation	Theory CIA: 12 Attendance: 3 Semester Exam: 45	Practical CA: 38 Attendance: 2
Paper Structure for Theory Semester Exam Module : Answer 3 out of 5 of 15 marks each		

Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
CO1	Recall and explain fundamental concepts of data science, data types, ML lifecycle, and statistics.	K1-K2 (Remember/Explain)
CO2	Apply data preprocessing techniques including cleaning, integration, transformation, and visualization.	K3 (Apply)
CO3	Analyze datasets using EDA, descriptive statistics, probability, correlation, and regression methods.	K4 (Analyse)
CO4	Design predictive models using regression, curve fitting, and feature engineering techniques.	K6 (Create)
CO5	Evaluate model performance using hypothesis testing and validation metrics.	K5(Evaluate)
CO6	Create end-to-end data science projects integrating preprocessing, modeling, and evaluation.	K6 (Create)