

**22CSE02 - DATA SCIENCE**

Programme & Branch	B.E. - Computer Science and Engineering	Sem.	Category	L	T	P	Credit
Prerequisites	NIL	5	PE	3	0	0	3
Preamble	This course integrates the fields within computer science, machine learning, statistics and hypothesis testing in the context of data science and its applications to create adept and well-rounded data scientist						
Unit - I	Introduction						9
Introduction– Data Science - Data Science and Other Fields - The Relationship between Data Science and Information Science - Computational Thinking - Issues of Ethics, Bias, and Privacy in Data Science - Data Types – Data Collections – Data Pre-processing. Techniques: Data Analysis and Data Analytics - Descriptive Analysis - Diagnostic Analytics - Predictive Analytics - Prescriptive Analytics - Exploratory Analysis - Mechanistic Analysis.							
Unit - II	Machine Learning						9
Introduction- Linear Regression – Multiple Linear Regression - Gradient Descent – Supervised Learning: kNN – Decision Tree – Naïve Bayes - Unsupervised Learning : k-means - Expectation Maximization – Reinforcement Learning							
Unit - III	Applications, Evaluations, and Methods						9
Solving Data Problems: Collecting and Analyzing Twitter Data – Collecting and Analyzing YouTube Data – Analyzing Yelp Reviews and Ratings. Data Collection Methods – Picking Data Collection and Analysis Method: Quantitative Methods, Qualitative Methods – Evaluation: Comparing Models – Cross-Validation.							
Unit - IV	Statistics						9
Role to Statistics -Estimation of Parameter and Sampling Distribution: Point Estimation - Sampling Distributions and the Central Limit Theorem. Statistical Intervals for a Single Sample: Confidence Interval on Mean – variance and Standard Deviation - Guidelines - Bootstrap - Tolerance and Prediction Intervals.							
Unit - V	Hypothesis Testing						9
Hypothesis Testing - Tests on the Mean, Variance and Standard Deviation of Single Sample and Two Samples - Nonparametric Test for Single Sample and Two Samples - Hypothesis Tests in Simple Linear Regression - Multiple Linear Regression.							
TEXT BOOK:							
1	Chirag Shah, “A Hands-On Introduction to Data Science”, 1st Edition, Kindle Edition, 2020 for Units I,II,III						
2	Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for Engineers, Sixth Edition, Wiley, 2013 for Units IV,V						
REFERENCES:							
1	Joel Grus, "Data Science from the Scratch", O'Reilly, 2015						
2	Frank Kane, “Hands-On Data Science and Python Machine Learning”, First edition, Packt Publication, 2017						



COURSE OUTCOMES: On completion of the course, the students will be able to												BT Mapped (Highest Level)			
CO1	apply preprocessing techniques to collect, clean, and prepare data and Visualize and present the inference using various tools												Applying (K3)		
CO2	use machine learning techniques to solve real time problems												Applying (K3)		
CO3	utilize the data analysis techniques for handling applications with large data												Applying (K3)		
CO4	make use of the statistical foundations of data science and analyze the degree of certainty of predictions using statistical test and models												Applying (K3)		
CO5	structure engineering decision making problem as hypothesis tests												Applying (K3)		
Mapping of COs with POs and PSOs															
COs/ POs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	
CO1	3	2	1	1					1	1		1	3	1	
CO2	3	2	1	1					1	1		1	3	1	
CO3	3	2	1	1					1	1		1	3	1	
CO4	3	2	1	1					1	1		1	3	1	
CO5	3	2	1	1					1	1		1	3	1	
1 – Slight, 2 – Moderate, 3 – Substantial, BT- Bloom's Taxonomy															
ASSESSMENT PATTERN – THEORY															
Test / Bloom's Category*	Remembering (K1) %		Understanding (K2) %		Applying (K3) %		Analyzing (K4) %		Evaluating (K5) %		Creating (K6) %		Total %		
CAT1	-		50		50								100		
CAT2	-		40		60								100		
CAT3	-		20		80								100		
ESE	-		40		60								100		
* ±3% may be varied (CAT 1, 2, 3 – 50 marks & ESE – 100 marks)															