

CSE11120	Introduction to Data Science	L	T	P	C
Version 1.0	Contact Hours 45	3	0	0	3
Pre-requisite/Exposure	---				
Co-requisite	NIL				

Course Objectives:

- This course will provide a foundation in the area of data science based on data curation and statistical analysis.

Course Outcomes:

On the completion of this course the student will be able to

CO1: Define Understand the limitations of data sets based on their contents and provenance
CO2: Understand data organization, management, preservation, and reuse

Knowledge of what statistical analysis techniques to choose, given particular demands of inference and available data

CO3: Explain of general linear models and cluster analysis methods for statistical analysis
CO4: Demonstrate Skills and knowledge in preparing data for analysis, including cleaning data, manipulating data, and dealing with missing data

CO5: Show Skills in actually analyzing data using open source data analysis tool

Course Description:

The primary goal of this course is for students to learn data analysis concepts and techniques that facilitate making decisions from a rich data set. Students will investigate data concepts, metadata creation and interpretation, general linear method, cluster analysis, and basics of information visualization. At the beginning, this course will introduce fundamentals about data and data standards and methods for organizing, curating, and preserving data for reuse. Then, we will focus on the inferential statistics: drawing conclusions and making decisions from data. This course will help students understand how to use data analysis tools, and especially, provide an opportunity to utilize an open source data analysis tool, R, for data manipulation, analysis, and visualization. Finally, in this course we will discuss diverse issues around data including technologies, behaviors, organizations, policies, and society.

Course Content:

Units	Lecture Hours
Unit-I: Definition of Data Science, Essential Concepts of Data, Data Problems and Solutions	5
Unit-II: Data Modelling and Relationship, Data Structure and Variables.	

Units	Lecture Hours
Unit-III: Descriptive Statistics, Sampling Distributions, Big Data and Statistics	
Unit-IV: Linear Regression, Cluster Analysis, Similarity Metrics, Dimensionality Reduction	
Unit-V: Text Analysis(Unstructured Data), Relevant Issues in Data Science	

Text Books:

- Jeffrey M. Stanton (2013). Introduction to Data Science.
- 10 Lecture Hours
- 10 Lecture Hours
- 10 Lecture Hours
- 10 Lecture Hours

Examination Scheme:

Components	Mid Term	Class Assessment	End Term
Weightage (%)	20	30	50

Relationship between Course Outcomes (COs) and Program Outcomes (POs):

1 = Weakly Mapped

2 = Moderately Mapped 3 = Strongly Mapped

Course Code	Course Name	COs	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PS O1	PS O2	PS O3
CSE11	Introduc	CO111	3	3	2	3	2	2	2	-	3	-	-	2	2	1	1
120	tion to	20.1															
	Data Science	CO111 20.2	3	2	2	2	2	2	2	-	3	-	-	1	3	3	3
		CO111	2	3	2	3	2	1	3	-	3	-	-	1	3	3	3
		20.3															
		CO111	2	2	3	3	2	2	1	-	3	-	-	2	2	2	2
		20.4															
		CO111	3	3	2	3	2	2	1	-	3	-	-	1	1	2	3
		20.5															
		CO111	2.	2.	2.	2.	2.	1.	1.	-	3.	-	-	1.4	2.2	2.2	2.4
		20	6	6	2	8	0	8	8		0						