CSE11153	Data Analysis	L	T	P	C
Version 1.0	Contact Hours 45 Hours	3	0	0	3
Pre-requisite/Exposure	Discrete Mathematics, Calculus, Machine Learning				
Co-requisite	NIL				

Course Objectives:

- To introduce fundamental concepts of data analysis
- To develop statistical models to represent data
- To provide intelligent solutions for data mining processes

Course Outcomes:

On the completion of this course the student will be able to CO1: Define different types of data analysis

techniquesCO2: Compare various statistical tests

CO3: Explain data using different types of data distributionsCO4: Build algorithms for data processing

CO5: Interpret different types of data visualization.

Course Description:

Data mining, or knowledge discovery in databases, has during the last few years emerged as one of the most exciting fields in Computer Science. Data mining aims at finding useful regularities in large data sets. Interest in the field is motivated by the growth of computerized data collections which are routinely kept by many organizations and commercial enterprises, and by the high potential value of patterns discovered in those collections. In this course we explore how this interdisciplinary field brings together techniques from databases, statistics, machine learning, and information retrieval. We will discuss the main data mining methods currently used, including data warehousing and data cleaning, clustering, classification, association rules mining, query flocks, text indexing and searching algorithms,how search engines rank pages, and recent techniques for web mining.

Course Content:

Units	Lecture Hours
Unit-I: Introduction to Data Analytics: Introduction to core concepts and technologies: Introduction, Terminology, data analysis process, types of data,Example applications Basic Review on Statistics I: Distributions, Estimation.	10
Unit-II: Probability & Statistics: Basic Review on Statistics II: Type I and II errors, rejection regions; Z-test, T-test, F-test, Chi-Square test, Bayesian test, Markov	11

Units	Lecture Hours
process, Hidden Markov Models, Poisson Process, Bayesian Network, Regression,Queuing systems.	
Unit-III: Business Process Understanding & Acquiring Data: Data Exploration: Observation, Interviews, Other sources. Data collection and management: Introduction, Sources of data, Data collection and APIs, Exploring and fixing data, Data storage and management, using multiple datasources	**12**
Unit-IV: Social Media and Text Analytics: Data preparation: Removing unwanted data, missing data handling Data Visualization: Introduction, Types of datavisualization, recent trends in various data collection and analysis techniques, Data Mining Trends and Research Frontiers: Mining Complex Data Types, Other Methodologies of Data, Mining, Data Mining Applications, Data Mining and Society, Data Mining Trends.	**12**

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 Mapped3 = 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
CSE111	Data	CO1115	3	2	3	2	1	1	1	-	3	-	-	1	3	2	3
53	Analy	3.1															
	sis	CO1115	2	2	2	2	3	3	3	-	3	-	-	3	2	1	3
		3.2															
		CO1115	2	3	3	2	2	2	2	-	2	-	-	3	2	1	2
		3.3															
		CO1115	1	2	3	3	3	3	1	-	3	-	-	3	2	3	2
		3.4															
		CO1115	1	1	2	2	3	3	2	-	2	-	-	2	2	3	1
		3.5															
		CO1115	1.	2.	2.	2.	2.	2.	1.	-	2.	-	-	2.4	2.2	2.0	2.2
		3	8	0	6	2	4	4	8		6						