

| | | | | |
|---|--|----------|----------|---|
| Data Science and Visualization | 3 | 0 | 2 | Probability, Statistics, Linear Algebra |
| | | | | |
| Course Objective: 1. To familiarize with different types of data and its visualization. 2.To understand and practice data pre-processing and data exploration. 3. To solve real-world analytical problems in data science. | | | | |
| | | | | |
| S. NO. | Course Outcomes (CO) | | | |
| CO1 | Ability to identify different types of data and data distributions. | | | |
| CO2 | Ability to understand and apply different data cleaning and data transformation techniques. | | | |
| CO3 | Ability to understand and implement different data visualization techniques. | | | |
| CO4 | Ability to understand and execute different data exploration techniques. | | | |
| CO5 | Ability to implement different real-world applications of data science. | | | |
| | | | | |
| S. NO. | Contents | | | Contact Hours |
| UNIT 1 | Introduction to data science: Basics of Probability & Statistics (Random Variables, Bayes's Theorem, Normal distribution, Central Limit Theorem). Defining data science, Recognizing different types of data, Data distributions. Data acquisition and data storage. | | | 10 |

| | | |
|---------------|--|-----------|
| UNIT 2 | Data pre-processing: Missing data problem, Outlier definition. Data cleaning, Data transformation or data wrangling procedures such as merging, ordering and aggregating. | 10 |
| UNIT 3 | Data visualization: Introduction to data visualization. Definition of Dashboard, Dashboard design and principles. Basic charts and plots, Box plots, Histogram, Graphs, Networks, Hierarchies, Reports. | 10 |
| UNIT 4 | Exploratory Data Analysis: Data exploration for univariate data. Outlier detection techniques. Descriptive statistics (mean, standard deviation etc.) for data exploration. Correlation statistics for data exploration. Data exploration for multivariate data. Use of multivariate visualization tools such as bar charts, bar plots, heat maps, bubble charts, run charts, and scatter plots. | 12 |
| | TOTAL | 42 |

REFERENCES

| S.No. | Name of Books/Authors/Publishers | Year of Publication / Reprint |
|--------------|---|--------------------------------------|
| 1 | Data science from scratch, Joel Grus, 2nd ed., O'Reilly Media. | 2019 |
| 2 | Doing data science: Straight talk from the frontline, Cathy O'Neil, Rachel Scutt, O'Reilly Media. | 2013 |
| 3 | Python data science handbook, Jake VanderPlas, 2nd ed., O'Reilly Media. | 2016 |
| 4 | Introducing data science, Davy Cielen, Arno D.B. Meysman, Mohamed Ali, 1st ed., Manning publications. | 2016 |
| 5 | The data science handbook, Field Cady, John Wiley & sons. | 2017 |

B.Tech. Information Technology

| Course code: Course Title | Course Structure | | | Pre-Requisite |
|----------------------------------|-------------------------|----------|----------|----------------------|
| | L | T | P | |