	Introduction to Data Science	L	T	P	C
		3	0	0	3
Pre-requisites/Exposure	--				
Co-requisites	--				

CO1 Gain a basic understanding of the role of Data Science in various scenarios in the real world of business, industry, and government and various roles and stages in a Data Science Project and ethical issues to be considered.
CO2 Explore the processes, tools, and technologies for the collection and analysis of structured and unstructured data.
CO3 Appreciate the importance of techniques like data visualization for the effective presentations of the outcomes with the stakeholders.
CO4 Understand techniques for preparing real-world data for data analytics.
CO5 Implement data analytic techniques for discovering interesting patterns from data.

Unit 1: Introduction to Data Science 4 Lecture hours
Fundamentals of Data Science, Real World Applications, Data Science Challenges, Data Science Teams, and Software Engineering for Data Science (DataOps, MLOps (intro)).

Defining Analytics, Types of data analytics (Descriptive, Diagnostic, Predictive, Prescriptive) Data Science Process: CRISP-DM Methodology, SEMMA, BIG DATA LIFE CYCLE, SMAM.

Unit 2: Data and Data Models 8 Lecture hours
Types of Data and Datasets, Data Quality, and Issues, Data Models, General Framework of Formal modeling: Association Analyses, Prediction Analyses, Data Pipelines and patterns, Data Pipeline Stages (extraction, ingestion, cleaning, exploration, wrangling, versioning, Data transformation, Feature management)

Modern Data Infrastructure: Diverse data sources, Cloud data warehouses, and lakes

Data wrangling (Data cleaning, Data Aggregation, Sampling, Statistical descriptions of data, measuring data similarity & dissimilarity), Handling Numeric Data (Discretization, Binarization, Normalization, Data Smoothening), Dealing with textual Data, Images, audio, and video data, Managing Categorical Attributes, Overview of visualization techniques for Data Exploratory analysis.

Unit 3: Feature Engineering 6 Lecture hours

Feature Extraction, Feature Construction, Feature Subset selection, Feature Learning, Feature Reduction (Dimensionality Reduction) Case Study involving FE tasks, and Feature Engineering techniques for text, images, audio, and video.

Unit 4: Classification, Prediction, and Association Analysis 10 Lecture hours
Concepts of classification and prediction, Decision trees for classification - ID3 algorithm

using entropy and Gini Index, Evaluation of classification algorithms, Association Analysis Association analysis concepts, Apriori Algorithm for frequent item sets, FP Growth for frequent item sets, Mining association rules

Unit 5: Clustering and Anomaly Detection **8 Lecture hours**

Cluster analysis concepts, Partitioning methods – k-Means algorithm, Hierarchical methods for cluster analysis, Density-based methods for cluster analysis – DBSCAN, Evaluation of clustering algorithms.
Anomaly Detection: Concepts of Outliers, Statistical approaches, Proximity, and Density-based outlier detection.

Modes of Evaluation: Quiz/Assignment/ presentation/ extempore/ Written Examination
Examination Scheme:

Components	Internal	Mid Term	ESE	Total
Weightage (%)	30%	20%	50%	100%

Relationship between the Program Outcomes (POs), Program Specific Outcomes, and Course Outcomes (COs)

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2
CO1												
CO2												
CO3												
CO4												
Average												

1. WEAK 2. MODERATE 3. STRONG