PVP23

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY

(Autonomous)

KANURU, VIJAYAWADA-520007

DEPARTMENT OF CSE (Data Science)

II B. Tech – I Sem CSE (DATA SCIENCE)

INTRODUCTION TO DATA SCIENCE

Course Code	23ES1306	**Year**	II	**Semester**	I
Course Category	Engineering Science	**Branch**	CSE (DATA SCIENCE)	**Course Type**	Theory
Credits	3	**L-T-P**	3-0-0	**Prerequisites**	Probability & Statistics
Continuous Internal Evaluation	30	**Semester End Examination**	70	**Total Marks**	100

Course Outcomes		
Upon successful completion of the course, the student will be able to:		
CO1	Describe the fundamental concepts and tasks involved in the Data Science life cycle.	**L2**
CO2	Apply data acquisition techniques to identify and collect data from various sources considering ethical and legal aspects.	**L3**
CO3	Apply exploratory data analysis techniques to summarize, visualize, and understand the relationships within a dataset.	**L3**
CO4	Analyze the effectiveness and limitations of different data acquisition methods, preprocessing techniques, and visualization tools in solving data science problems.	**L4**

Contribution of Course Outcomes towards achievement of Program Outcomes& Strength of correlations (3: High,2: Medium, 1: Low)														
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**
CO1	**2**													
CO2	**3**												**3**	
CO3	**3**												**3**	
CO4		**3**										**2**		

Unit No.	Syllabus Contents	Mapped CO
I	**Major Tasks in Data Science**: Data Collection, storing data, Data Processing, Exploratory Data Analysis, Data Modeling. **Life cycle of Data Science**: Business Understanding, Data Understanding, Data Preparation, Model Building, Model Evaluation, and Deployment, CRISP DM methodology **Applications of data science:** Finance, Healthcare, Business and Marketing, Manufacturing, Cyber security, Transportation, Social Media, Agriculture, etc	**CO1**
II	**Introduction to Data Acquisition:** Definition and importance in data science, Overview of data acquisition process, Types of data: structured, semi-structured, and unstructured **Data Sources**: Data Sources, Public datasets, Private datasets, APIs, Web scraping, Databases **Data Collection Methods**: Surveys, Experiments, Sensor data, Social media data, Transactional data **Data Formats**: CSV, JSON, XML **Ethical and legal considerations**: Licensing, copyright, consent, Anonymization and ethical considerations	**CO1, CO2, CO4**

III	**Data Preprocessing**: Introduction, Data Objects, Attribute Types-Nominal, Binary, Interval-Scaled, Ratio-Scaled, Discrete versus Continuous Attributes, Need of Data Preprocessing, Data Quality, Major Tasks in Data Preprocessing **Data Cleaning**: Missing Values, Noisy data, **Data Integration**: Entity Identification Problem, Redundancy and Correlation Analysis, Data Value Conflict Detection and Resolution **Data Transformation:** Smoothing, Normalization (Minmax, z-score, Decimal scaling) **Data Reduction:** Attribute Subset Selection, Sampling	**CO1, CO2, CO4**
IV	**Exploratory Data Analysis (EDA):** **Descriptive Statistics:** Measures of Central Tendency (Mean, median, mode), Measures of Dispersion (Range, variance, standard deviation, interquartile range), Distribution Shapes (Skewness, kurtosis), Data Summarization (Frequency tables, cross-tabulation.) **Correlation and Relationships:** Pearson Correlation Coefficient, Spearman and Kendall Correlation Coefficients	**CO1, CO3, CO4**
V	**Data Visualization Techniques:** Importance and Benefits of Data Visualization, Line Plots, Bar Charts, Histograms, Box Plots, Scatter Plots, Heatmaps, Violin Plots, Pair Plots	**CO1, CO3, CO4**

Learning Resources
Text Books
1) Introducing Data Science, David Cielen, Arno D. B. Meysman, and Mohamed Ali, 2016, Manning Publications. 2) Data Science: Concepts and Practice, Vijay Kotu, Bala Deshpande, Second Edition, 2018, Morgan Kaufmann. 3) Data Mining: Concepts and Techniques, Jiawei Han, Micheline Kamber and Jian Pei, Third edition, Morgan Kaufmann.
References
1) Data Science from Scratch: First Principles with Python, Joel Grus, Second edition, 2019, O'Reilly 2) Data Science for Business, Foster Provost, Tom Fawcett, First Edition, 2013, O'Reilly Media 3) Python for Data Analysis, Wes McKinney, Second Edition, 2017, O'Reilly Media
E-Resources
1. https://archive.nptel.ac.in/courses/106/106/106106179/ 2. https://www.youtube.com/watch?v=-ETQ97mXXF0 3. https://www.datacamp.com/tracks/associate-data-scientist-in-python