

PRASAD V. POTLURI SIDDHARTHA INSTITUTE OF TECHNOLOGY  
KANURU, VIJAYAWADA  
II B.Tech – I Sem, CSE(Data Science)

**Foundations of Data Science**

**Syllabus**

Offering Branches	CSE(Data Science)	Course Code:	20ES1308
Course Category:	Engineering Science	Credits:	3
Course Type:	Theory	Lecture-Tutorial-Practical:	3-0-0
Prerequisites:	Probability & Statistics	Continuous Internal Evaluation:	30
		Semester End Examinations	70
		Total Marks:	100

**COURSE OUTCOMES**

Upon successful completion of the course, the student will be able to:

<b>CO1</b>	Understand the life cycle process of data science.	<b>L2</b>
<b>CO2</b>	Apply different data pre-processing techniques to improve data quality.	<b>L3</b>
<b>CO3</b>	Apply different exploratory data analysis techniques to understand the relationship of data objects.	<b>L3</b>
<b>CO4</b>	Apply Linear and Classification methods for model building	<b>L3</b>

**Contribution of Course Outcomes towards achievement of Program Outcome & Strength of correlation (3: High, 2: Medium, 1:Low)**

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
<b>CO1</b>	3													
<b>CO2</b>	2													
<b>CO3</b>	2													
<b>CO4</b>	3											1		

<b>Course Contents</b>		
<b>UNIT-1</b>	<p><b>Introduction to Data Science:</b></p> <p><b>Major Tasks in Data Science:</b> Data Collection, storing data, Data Processing, Exploratory Data Analysis, Data Modeling.</p> <p><b>Life cycle of Data Science:</b> Business Understanding, Data Understanding, Data Preparation, Model Building, Model Evaluation, and Deployment.</p> <p><b>Applications of data science:</b> Finance, Healthcare, Business and Marketing, Manufacturing, Cyber security, Transportation, Social Media, Agriculture, etc</p>	<b>CO1</b>
<b>UNIT-2</b>	<p><b>Data Preprocessing:</b> Introduction, Types of Attributes, Need of Data Preprocessing, Data Quality, Major Tasks in Data Preprocessing</p> <p><b>Data Cleaning:</b> Missing Values, Noisy data,</p> <p><b>Data Integration:</b> Entity Identification Problem, Redundancy and Correlation Analysis, Data Value Conflict Detection and Resolution</p> <p><b>Data Transformation:</b> Smoothing, Normalization (Min-max, z-score, Decimal scaling)</p> <p><b>Data Reduction:</b> Attribute Subset Selection, Sampling</p>	<b>CO1, CO2</b>
<b>UNIT-3</b>	<p><b>Exploratory Data Analysis (EDA):</b></p> <p><b>Descriptive Statistics:</b> Measures of Central Tendency: Mean, Median, Mode, Measures of Variability or Dispersion: Range, Variance, Standard Deviation</p> <p><b>Data Visualization Techniques:</b> Univariate visualization: Histograms, box plots, bar plots, Bivariate visualization: Scatter plots, line plots. Multivariate visualization: Heatmaps, pair plots.</p> <p><b>Correlation and Relationships:</b> Pearson Correlation Coefficient, Spearman and Kendall Correlation Coefficients</p>	<b>CO1, CO3</b>
<b>UNIT-4</b>	<p><b>Linear methods for Regression:</b></p> <p>Introduction, Linear Regression models, Least Squares, Multiple Regression, examples</p> <p>Linear methods for Classification: Introduction, Logistic Regression, examples</p>	<b>CO1, CO4</b>
<b>UNIT-5</b>	<p><b>Classification Models:</b> Introduction, What is Classification, General Approach to Classification, Decision Tree Induction algorithm, examples, Model evaluation metrics.</p> <p><b>Bayes Classification Methods:</b> Bayes' Theorem, Naive Bayesian Classification algorithm, examples</p>	<b>CO1, CO4</b>

<b>Learning Resources</b>
<b>Text Books</b>
<ol style="list-style-type: none"><li>1. Introducing Data Science, David Cielen, Arno D. B. Meysman, and Mohamed Ali, 2016, Manning Publications. (UNIT-I)</li><li>2. Data Mining: Concepts and Techniques, Jiawei Han, Micheline Kamber and Jian Pei, Third edition, Morgan Kaufmann. (UNIT-II, III and V)</li><li>3. The Elements of Statistical Learning, Trevor Hastie, Robert Tibshirani, Jerome Friedman, Second Edition, Springer. (UNIT- IV)</li></ol>
<b>References</b>
<ol style="list-style-type: none"><li>1. Cathy O'Neil and Rachel Schutt, "Doing Data Science", O'Reilly, 2015.</li><li>2. Data Science from Scratch: First Principles with Python, Joel Grus, Second edition, 2019, O'Reilly</li><li>3. Statistics, Robert S. Witte and John S. Witte, Eleventh Edition, 2017, Wiley Publications.</li></ol>
<b>e- Resources &amp; Other digital material</b>
<ol style="list-style-type: none"><li>1. <a href="https://nptel.ac.in/courses/106106212">https://nptel.ac.in/courses/106106212</a></li><li>2. <a href="https://nptel.ac.in/courses/106106179">https://nptel.ac.in/courses/106106179</a></li><li>3. Data Science Methodology- Coursera - <a href="https://www.coursera.org/learn/datascience-methodology">https://www.coursera.org/learn/datascience-methodology</a></li><li>4. Foundations of Data Science - edX - <a href="https://www.edx.org/course/foundationsof-data-science">https://www.edx.org/course/foundationsof-data-science</a></li></ol>