



GUJARAT TECHNOLOGICAL UNIVERSITY

Minor Degree: Artificial Intelligence and Machine Learning

Subject Code: 115AG01

Semester – V

Subject Name: Introduction to Data Analytics

Prerequisite: Basic Programming

Rationale:

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks				Total Marks
L	T	P	C	Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	0	2	4	70	0	30	0	100

Content:

Unit No	Course Content	No of Hours
1	Introduction to Data Science, Different Sectors using Data science, Purpose and Components of Python in Data Science.	7
2	Data Analytics Process, Knowledge Check, Exploratory Data Analysis (EDA), EDA- Quantitative technique, EDA- Graphical Technique, Data Analytics Conclusion and Predictions.	7
3	Feature Generation and Feature Selection (Extracting Meaning from Data)- Motivating application: user (customer) retention- Feature Generation (brainstorming, role of domain expertise, and place for imagination)- Feature Selection algorithms.	11
4	Data Visualization- Basic principles, ideas and tools for data visualization, Examples of inspiring (industry) projects- Exercise: create your own visualization of a complex dataset.	10
5	Applications of Data Science, Data Science and Ethical Issues- Discussions on privacy, security, ethics- A look back at Data Science- Next-generation data scientists.	7
Total Hrs.		42

Suggested Specification table (Theory):

Distribution of Theory Marks (%)					
R Level	U Level	A Level	N Level	E Level	C Level

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)



GUJARAT TECHNOLOGICAL UNIVERSITY

Minor Degree: Artificial Intelligence and Machine Learning
Subject Code: 115AG01

Reference Books:

1. Joel Grus, Data Science from Scratch, Shroff Publisher Publisher /O'Reilly Publisher Media
2. Annalyn Ng, Kenneth Soo, Numsense! Data Science for the Layman, Shroff Publisher Publisher
3. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly Publisher Media.
4. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press.
5. Jake VanderPlas, Python Data Science Handbook, Shroff Publisher Publisher /O'Reilly Publisher Media
6. Philipp Janert, Data Analysis with Open Source Tools, Shroff Publisher Publisher /O'Reilly Publisher Media.

Course Outcomes: Upon completion of this course students should be able to:

No	Course Outcomes	% weightage
01	Explain how data is collected, managed and stored for data science;	
02	Understand the key concepts in data science, including their real-world applications and the toolkit used by data scientists;	
03	Implement data collection and management scripts using MongoDB	

List of Practical:

1. Python Environment setup and Essentials.
2. Mathematical computing with Python (NumPy).
3. Scientific Computing with Python (SciPy).
4. Data Manipulation with Pandas.
5. Prediction using Scikit-Learn
6. Data Visualization in python using matplotlib