GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Engineering

Level: PG

Branch: Internet of Things

Course / Subject Code : ME01062081

Course / Subject Name: Fundamentals of Data Science

w. e. f. Academic Year:	2024-25
Semester:	1st Semester
Category of the Course:	PEC – II

Prerequisite:	• Fundamental Knowledge of programming languages. • Advanced understanding of mathematics. • Aptitude for probabilities and statistics and training area like artificial intelligence
Rationale:	• The data Science brings together the domain expertise from programming, mathematics, and statistics to create insights and make sense of data. Making sense of data will reduce the horrors of uncertainty for organizations. There are different tools and techniques related to Data Science that are widely used in a variety of applications such as statistical analytics, time-series analytics. The preprocessing of the data is very important before any statistical or machine learning models are developed from it. This course covers the basics of data science, preprocessing, visualization, data modeling by statistical and machine learning methods.

Course Scheme:

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
				Theory		Practical		
L	T	PR	C	ESE (E)	PA(M)	ESE (V)	PA (I)	
03	00	02	4	70	30	30	20	150

Course content

Unit	Course Content	No of Hours	% of Weightage

1	**Fundamentals of Data Science:** Introduction to data science, Data analytics life cycle, Type of data analysis, Types of jobs in data analytics, Data science tools, Fundamental areas of study in data science, Role of SQL in data science, Pros and cons of data science	06	15%
2	**Data Preprocessing, Plotting and Visualization:** Data types and forms, Possible data error types, Various data preprocessing operations, Introduction to data visualization, Visual encoding, Data visualization libraries, Basic data visualization tools	8	20%
3	**Statistical Data Analysis and Machine Learning:** Role of statistics in data science, Descriptive statistics, Inferential statistics, Overview of machine learning, Supervised machine learning, Regression methods, Classification methods, Unsupervised machine learning, Clustering methods, Association analysis	14	30%
4	**Time-series Analysis:** Overview of time-series analysis, Components of time-series, Time-series forecasting models	8	20%
5	**Recent Trends in the Domain of Data Science**	6	15%

Textbook:

1. Dat Data Science Science Fundamentals and Practical Approaches by Gypsi Nandi and Rupam Kumar Sharma First Edition, BPB Publications India.

Reference Books:

1. Introducing Data Science by Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Latest Edition, Manning.
2. Data Science and Analytics with Python by Jesus Rogel- Salazar Latest Edition, CRC Press Taylor & Francis Group

Course Outcome:

After completion of the Course, Students will be able to:

No	Course Outcomes	RBT Level*

01	Understand the data analysis life cycle and concepts of data preprocessing, visualization, statistical and machine learning methods.	UN
02	Apply data preprocessing and data visualization methods on real world data.	AP
03	Apply programming languages and machine learning methods to develop predictive model from clean data.	AP
04	Analyze the working methodology of machine learning methods.	AN
05	Evaluate the performance of statistical and machine learning methods using Suitable performance metrics.	EV

*RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create

Suggested Course Practical List:

Minimum 10 practicals based on course content are to be performed.

List of Laboratory/Learning Resources Required:

1. List of Hardware

   Computer systems with latest configuration and connected in a LAN.
2. List of Software

   Python Programming Language and IDE, Open source python libraries for data preprocessing, visualization, machine learning methods, time-series analysis.
3. List of Useful websites/MOOCs

   Learners are advised to opt for NPTEL and SWAYAM courses that are relevant to this course.

****