GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering
Level: UG
Subject Code: BE05016021
Subject Name: Data Science

w.e.f.AcademicYear:	2024-25
Semester:	5
Category of the Course:	Professional Core Course

Prerequisite:	None
Rationale:	To get the knowledge of the mathematical foundations and methods needed for data science, Analyze data to make quicker and better decisions. Data science helps in managing, analyzing and understanding trends in data leading to design the strategy for better and sustainable environment around us

Course Outcomes:

Sr.No.	CO statement	Marks% weightage
CO-1	Understanding data science and analytics and their application in various domains	15
CO-2	Identify the data types, relation between data and visualization technique for data	15
CO-3	Explain probability, distribution, sampling, Estimation	25
CO-4	Solve regression and classification problem	25
CO-5	Explore the applications of data science through various case studies.	20

Teaching and Examination Scheme:

Teaching - Learning Scheme (in Hours per Semester)					Total Credits = TH/30	Assessment Pattern and Marks					Total Marks
						Theory		Tutorial / Practical			
L	T	P	PBL*	TH		ESE (E)	PA (M)	PA (I)	PBL (I)	ESE (V)	
45	0	30	15	90	03	70	30	20	30	50	200

Content:

Sr.No.	Content	Total Hrs
1	**Introduction to Data Science & Engineering:** Concept of Data Science; Data engineering vs. Analytics; Reporting vs. Analytics; Introduction to analytics process; Types of Analytical Techniques (Descriptive, Diagnostic, Predictive, Perspective); Traits of Big Data; Web Scrapping and Social Media Analytics	04

2	**Python Essentials for Data Science:** Introduction to Python programming; Data structures (Lists, Dictionaries, Sets); Vectorized operations with NumPy; Data manipulation and cleaning with Pandas (DataFrames, Series); Overview of the Matplotlib and Seaborn libraries for visualization.	**07**
3	**Descriptive Analytics & Visualization:** Data Types and Scales (Measurement Scales); Population vs. Sample; Percentile, Decile, and Quartile; Measures of Variation and Shape (Skewness and Kurtosis); Visualization techniques for different data relations.	**04**
4	**Probability Recap:** Brief review of Probability Theory Axioms and Random Variables; Summary of PDF and CDF; Overview of key distributions (Binomial, Poisson, Normal, Chi-Square)	**03**
5	**Sampling:** Introduction to Sampling, Population Parameters and Sample Statistic, Sampling, Probabilistic Sampling, Non-Probability Sampling, Sampling Distribution, Central Limit Theorem (CLT), Sample Size **Estimation:** Classical Methods of Estimation. Estimating the Mean, Standard Error of a Point Estimate, Prediction Intervals, Tolerance Limits, Estimating the Variance, Estimating a Proportion for single mean, Difference between Two Means, between Two Proportions for Two Samples and Maximum Likelihood Estimation.	**07**
6	**Simple Linear Regression and Correlation:** Introduction to Linear Regression, The Simple Linear Regression Model, Least Squares and the Fitted Model, Properties of the Least Squares Estimators, Inferences Concerning the Regression Coefficients, Prediction, Simple Linear Regression Case Study	**05**
7	**Logistic Regression:** Introduction – Classification Problems, Introduction to Binary Logistic Regression, Estimation of Parameters in Logistic Regression, Interpretation of Logistic Regression Parameters, Logistic Regression Model Diagnostics, Classification Table, Sensitivity, and Specificity, Optimal Cut-Off Probability, Variable Selection in Logistic Regression, Application of Logistic Regression in Credit Rating, Gain Chart and Lift Chart	**05**
8	**Classification:** Overview of a Decision Tree, Introduction Chi-Square Automatic Interaction Detection (CHAID), The General Algorithm, Decision Cost-Based Splitting Criteria, Tree Algorithms – ID3, Naïve Bayes, Bayes‘ Theorem, Naïve Bayes Classifier, Evaluating a Decision Tree, Classification and Regression Tree Ensemble Method, Random Forest	**06**
9	**Applying Domain Expertise to Solve Real-World Problems Using Data Science:** Case Study 1: Data Science in healthcare Case Study 2: Delving into customer personality analysis Case Study 3: Data Science for Driving Growth in E-Commerce Case Study 4: Data Science in sentiment analysis	**04**
	TOTAL	**45**

Suggested Specification table with Marks(Theory): (ForB.E.only)

Distribution ofTheory Marks					
RLevel	ULevel	ALevel	NLevel	ELevel	CLevel
20	**25**	**25**	**--**	**--**	**--**

R:Remembrance;U:Understanding;A:Application,N:AnalyzeandE:EvaluateC:Create and above Levels (Revised Bloom's Taxonomy)

Note:This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Reference Books:

1. Business Analytics, Dinesh Kumar, Wiley India
2. Fundamentals of Data Science, Dr. B Dwarkanath, R M Rani, Dr. D. Usha, Notion Press
3. Data Science & Analytics, V K Jain, Khanna Book Publishing, New Delhi
4. Data Analytics, Anil Maheshwari, McGrawHill
5. Practical Statistics for Data Scientists, Peter Bruce and Andrew Bruce, O'Reilly
6. Introduction to Probability, Joseph K. Blitzstein and Jessica Hwang, CRC Press

List of Open Source Software/learning website:

1. **https://nptel.ac.in/courses/106106179**

2. **https://www.geeksforgeeks.org/machine-learning/supervised-machine-learning**

- **List of suggested activities for Problem Based Learning:**

Sl. No.	Name of the activity	No. of hours	Evaluation Criteria
1	Assignment writing. Numerical based assignment is preferable.	5 assignments of 3h each. Total = 15h	Based on the assignment submitted.
2	Problem solving/Coding using C, C++, Python, SCILAB, MATLAB, MS-EXCEL or any other relevant software	5 small coding-based problems of 3h each. Total = 15h	Based on the coding solution submitted.

3	Technical Video based learning related to the subject	Duration of video = 5h Report preparation & Presentation = 10h Total = 15h	Report /presentation based on the video learning outcomes.
4	Discussion on research paper based on relevant subject	3 research paper = 15h	Summarize research paper and evaluation critical parameters
5	Poster/chart/power point preparation on technical topics	Duration = 10 h	Based on poster/chart preparation and presentation skills
6	Application/Software development	Duration = 15 h	Depending on the complexity of the Application/Software
7	Group Discussion on emerging/trending technical topics based on subject	Duration = 1 h each	Based on performance in group discussion, technical depth, knowledge etc.
8	Seminar / Presentation	Duration for study and preparation=5h Report writing=3h Presentation=2h Total=10h	Topic can be selected technical content beyond syllabus
9	Real world case studies-based learning	Duration of data collection/study = 5h Report preparation = 10h Total = 15h	Based on in-depth study, technical depth, data collected, fact finding, etc.
10	Working/non-working model on technical topics	Working = 12 h Non- working = 8 h	Based on inter department/external evaluation
11	Self-learning on-line course	Minimum duration of the course should be 15h.	Examination based assessment at the end of course. Based on the certificate produced.
12	Complex problem solving	Maximum 3 problem. Study of the problem and solution finding, Total = 15h	Based on the depth of the solution submitted.
13	Industry/Research laboratory visit	Visit = 5h, Report preparation = 5h Total = 10h	Based on report submitted. Report should contain observations and calculations based on industry/ lab data.
14	Videos on Industrial safety aspects based on subject	Duration of video = 5h Report preparation = 5h Total = 10h	Based on quiz/report submitted
15	Industrial exposure for 2-3 days to observe and provide tentative solutions on society/environment /health/any other issue	Duration = 15 h for industrial exposure Problem identification and tentative solution = 10 h Total = 20 h	Based on evaluation of critical problems and solutions

Note

- All the suggested activity should be related to the subject.
- Min 3 activities must be carried out as per the availability of faculties and students.
- The number of hours is suggestive. Faculty can sub-divide the number of hours based on the activity. However, total number of hours is fixed.
- Rubrics for the evaluation can be prepared by the faculty.