



# GUJARAT TECHNOLOGICAL UNIVERSITY

**Bachelor of Engineering**

**Subject Code: 3144301**

**Semester – IV**

**Subject Name: Data Science for Humanity**

**Prerequisite:** None

**Rationale:** To get the knowledge of the mathematical foundations needed for data science, Analyzed 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.

## Teaching and Examination Scheme:

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

## Contents:

Sr. No.	Content	Total Hrs
1	<b>Introduction to Data Science:</b> Concept of Data Science, What is data engineering.?, Analytics vs. Reporting, Introduction to analytics process, Types of Analytical Techniques in BI – Descriptive, Predictive, Perspective, Social Media Analytics, Behavioral, Iris Datasets, Traits of Big Data, Web and Social Media Analytics, Web Scrapping, Machine Learning Algorithms Framework for Data-Driven Decision Making Analytics Capability Building Roadmap for Analytics Capability Building Challenges in Data-Driven Decision Making and Future	05
2	<b>Descriptive Analytics:</b> Introduction to Descriptive Analytics, Data Types and Scales, Types of Data Measurement Scales, Population and Sample, Percentile, Decile and Quartile, Measures of Variation, Measures of Shape – Skewness and Kurtosis	05
3	<b>Introduction to Probability:</b> Introduction to Probability Theory, Probability Theory – Terminology, Fundamental Concepts in Probability – Axioms of Probability, Random Variables, Probability Density Function (PDF) and Cumulative Distribution Function (CDF) of a Continuous Random Variable, Binomial Distribution, Poisson Distribution, Geometric Distribution, Parameters of Continuous Distributions, Uniform Distribution, Exponential Distribution, Chi-Square Distribution, Student's t-Distribution, F-Distribution	07
4	<b>Sampling:</b> Introduction to Sampling, Population Parameters and Sample Statistic, Sampling, Probabilistic Sampling, Non-Probability Sampling, Sampling Distribution,	08



# GUJARAT TECHNOLOGICAL UNIVERSITY

## Bachelor of Engineering

Subject Code: 3144301

	Central Limit Theorem (CLT), Sample Size  <b>Estimation:</b> 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.	
5	<b>Simple Linear Regression and Correlation:</b> 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	04
6	<b>Logistic Regression:</b> 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
7	<b>Classification:</b> 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
8	<b>Case Studies:</b> Healthcare Analytics, Data Science in BioTech, Data Science in Education, Weather Forecasting, Object Recognition, Customer Personality Analysis, Real Time Sentiment Analysis.	06
<b>Total</b>		<b>46</b>

### Suggested Specification table\* with Marks (Theory):

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
<b>20</b>	<b>25</b>	<b>25</b>	---	---	---

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



# GUJARAT TECHNOLOGICAL UNIVERSITY

## Bachelor of Engineering

**Subject Code: 3144301**

*\*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 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

**Course Outcomes:** Students will be able to

Sr. No.	CO Statement	Weightage (%)
1	Understanding data science and analytics and their application in various domains	10
2	Identify the data types, relation between data and visualization technique for data	15
3	Explain probability, distribution, sampling, Estimation	25
4	Solve regression and classification problem.	25
5	Explore the applications of data science through various case studies.	25