FACULTY OF ENGINEERING & TECHNOLOGY

First Year Master of Technology

Semester I/II

Course Code: 102341202

Course Title: Data Science

Type of Course: Program Elective VI/Core Course II

Course Objectives: To learn basic concepts of data collection, storage, processing, and visualization. To study and apply various methods of Descriptive Analysis. To study and investigate methods of data analysis and predictive modelling.

Teaching & Examination Scheme:

Contact hours per week			Course Credits	Examination Marks (Maximum / Passing)				
Lecture	Tutorial	Practical		Internal		External		Total
				Theory	J/V/P*	Theory	J/V/P*	
3	0	2	4	30 / 15	20 / 10	70 / 35	30 / 15	150 / 75

* **J**: Jury; **V**: Viva; **P**: Practical

Detailed Syllabus:

Sr.	Contents	Hours
1	Introduction to Data Science: A Brief History of Data Science, A History of Data Gathering, A History of Data Analysis, The Emergence and Evolution of Data Science, Where Is Data Science Used?, Data Science Use Cases, Myths about Data Science, Recent Trends.	4
2	Descriptive analytics: Introduction, Data types and Scale, measurement scales, population and sample, measure of central tendency, Percentile, decile and Quartile, Variation, measure of Shape	5
3	Introduction to Probability: Introduction to Probability Theory, Probability Theory – Terminology, Fundamental Concepts in Probability – Axioms of Probability, Application of Simple Probability Rules – Association Rule Learning, Bayes' Theorem, 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	6

4	Sampling and Estimation: Introduction to Sampling, Population Parameters and Sample Statistic, Sampling, Probabilistic Sampling, Non-Probability Sampling, Sampling Distribution, Central Limit Theorem (CLT), Sample Size Estimation for Mean of the Population, Estimation of Population Parameters, Method of Moments, Estimation of Parameters Using Method of Moments, Estimation of Parameters Using Maximum Likelihood Estimation	5
5	Correlation Analysis: Introduction, Person Correlation Coefficient, Spearman Rank Correlation, Point Bi-Serial Correlation, The Phi-coefficient	4
6	Simple Linear Regression : Introduction to Simple Linear Regression, History of Regression–Francis Galton's Regression Model, Simple Linear Regression Model Building, Estimation of Parameters Using Ordinary Least Squares, Interpretation of Simple Linear Regression Coefficients, Validation of the Simple Linear Regression Model, Outlier Analysis, Confidence Interval for Regression Coefficients b0 and b Confidence Interval for the Expected Value of Y for a Given X, Prediction Interval for the Value of Y for a Given X	6
7	Decision Trees: Introduction, Chi-Square Automatic Interaction, Detection (CHAID), Classification and Regression Tree, Cost-Based Splitting Criteria Ensemble Method, Random Forest	4
8	Introduction to Visualization: What Is Visualization? Visualization Foundation, Storytelling with Data, Visualization techniques for Different data , case study	6

Suggested Specification table with Marks (Theory) (Revised Bloom's Taxonomy):

Distribution of Theory Marks						R: Remembering; U: Understanding; A: Application, N: Analyze; E: Evaluate; C: Create
R	**U**	**A**	**N**	**E**	**C**	
10%	20%	25%	30%	10%	5%	

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: The science of Data Driven Decision Making by u Dinesh Kumar, Willey
2	Essential of Business Analytics by by Jeffrey D. Camm, James J. Cochran, Michael J. Fry, Jeffrey W. Ohlman, David R. Anderson
3	Business analytics Principles, Concepts, and Applications by Marc J. Schniederjans, Dara G. Schniederjans, Christopher M. Starkey, Pearson FT Press
4	Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic, Wiley

Course Outcomes (CO):

Sr.	Course Outcome Statements	%weightage
CO-1	Understand and apply concepts of probability and statistics in data analysis	**40**
CO-2	Understand the concept of data presentation of knowledge.	**20**
CO-3	Use data analysis and visualization tools to carry out basic statistical modeling and analysis.	**40**

List of Practicals / Tutorials:

1	Perform Descriptive statistics on given data sets
2	Preparing your data for analysis (preprocessing)
3	Apply Probability models for data set
4	Use various Sampling Techniques for data
5	Apply Regression model for prediction
6	Explore Classification methods on different data sets
7	Develop Various Charts and graphs using visualization tool (Bar chart, Line Chart, Pie Chart, Scatter Plot etc..)

Supplementary learning Material:

1	Open source Tools (R Studio, Python, Candela, etc.)
2	NPTEL Data Science for Engineers https://nptel.ac.in/courses/106/106/106106179/
3	towardsdatascience.com
4	r4ds.had.co.nz

Curriculum Revision:

Version:	**1**
Drafted on (Month-Year):	Apr-20
Last Reviewed on (Month-Year):	Jul-20
Next Review on (Month-Year):	Apr-22