FACULTY OF ENGINEERING & TECHNOLOGY

Effective from Academic Batch: 2022-23

Programme: **Bachelor of Technology (Artificial Intelligence (AI) and Data Science)**

Semester: **VI**

Course Code: **202046717**

Course Title: **Data Science**

Course Group: **Professional Core Course**

Course Objectives: To learn basic concepts of data collection, storage, processing, and its application. 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)				
				Theory		J/V/P*		Total
Lecture	Tutorial	Practical		Internal	External	Internal	External	
3	0	2	4	50 / 18	50 / 17	25 / 09	25 / 09	150 / 53

* **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.	04
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	05
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	07

4	**Correlation Analysis:** Introduction, Person Correlation Coefficient, Spearman Rank Correlation, Point Bi-Serial Correlation, The Phi-coefficient	**04**
5	**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	**06**
6	**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	**06**
7	**Decision Trees:** Introduction, Chi-Square Automatic Interaction, Detection (CHAID), Classification and Regression Tree, Cost-Based Splitting Criteria Ensemble Method, Random Forest	**05**
8	**Case study on Data science application:** Government, Sports, Banking, Healthcare, etc.	**03**
	Total	**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	Find correlation between gender and Semester marks.
5	Apply Regression model for prediction.
6	Prepare dataset and apply simple linear regression also visualize outlier from datasets.
7	Perform logistic regression on given data set.
8	Study and perform decision tree.
9	Prepare dataset and apply random forest algorithm on dataset.
10	Implement chi-squared algorithm.
11	Case study on any real time application.

Reference Books:

1	Business analytics: The science of Data Driven Decision Making by u Dinesh Kumar, Willey
2	Data science for Dummies by Lillian Pierson WILEY publication
3	Data Analytics for Beginners: Basic Guide to Master Data Analytics Paperback –by Paul Kinley
4	Doing Data Science by Cathy O'Neil, Rachel Schutt , O'Reilly Media, Inc.

Supplementary learning Material:

1	Lecture Note
2	NPTEL
3	https://www.analyticsvidhya.com/
4	Coursera

Pedagogy:

- Direct classroom teaching
- Audio Visual presentations/demonstrations
- Assignments/Quiz
- Continuous assessment
- Interactive methods
- Seminar/Poster Presentation
- Industrial/ Field visits
- Course Projects

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

Distribution of Theory Marks in %						**R**: Remembering; **U**: Understanding; **A**: Applying; **N**: Analyzing; **E**: Evaluating; **C**: Creating
R	**U**	**A**	**N**	**E**	**C**	
15%	20%	25%	30%	10%	---	

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.

Course Outcomes (CO):

Sr.	Course Outcome Statements	%weightage
CO-1	Identify the data types, the relationship between data, and processing method for data.	**20**
CO-2	Understand and apply concepts of descriptive analytics in data science.	**25**
CO-3	Understand the concept of regression for data analysis.	**25**
CO-4	To be able to apply the concept of different algorithms and methods on real time application.	**30**

Curriculum Revision:	
Version:	2.0
Drafted on (Month-Year):	June-2022
Last Reviewed on (Month-Year):	-
Next Review on (Month-Year):	June-2026