

गोंय विद्यापीठ

ताळगांव पठार,

गोंय - ४०३ २०६

फोन : + ९१ - ८६६९६०९०४८



Goa University

Taleigao Plateau, Goa - 403 206

Tel : +91-8669609048

Email : registrar@unigoa.ac.in

Website : www.unigoa.ac.in

(Accredited by NAAC with Grade A+)

GU/Acad –PG/BoS -NEP/2024/163

Date: 23.06.2026

CIRCULAR

In supersession to the Circular No. GU/Acad –PG/BoS -NEP/2024/844 dated 27.03.2026, the syllabus for the Semester V and VI of the **Master of Science (Integrated) in Data Science** Programme approved by the Standing Committee of the Academic Council in its meeting held on 24th February 2026 is enclosed. Further, the syllabus for Semester I to IV which was approved earlier is also attached.

The Dean/ Vice-Deans of the Goa Business School are requested to take note of the above and bring the contents of the Circular to the notice of all concerned.

(Ashwin V. Lawande)
Deputy Registrar – Academic

To,

1. The Dean, Goa Business School, Goa University.
2. The Vice-Deans, Goa Business School, Goa University.

Copy to:

1. The Chairperson, BOS in Data Science and Artificial Intelligence.
2. The Controller of Examinations, Goa University.
3. The Assistant Registrar, UG Examinations, Goa University.
4. The Assistant Registrar, PG Examinations, Goa University
5. Directorate of Internal Quality Assurance, Goa University for uploading the Syllabus on the University website.

Programme Structure for Semester I to X – Master of Science (Integrated) in Data Science										
Semester	Major -Core	Minor/VET	MC	AEC	SEC	I	D	VAC	Total Credits	Exit
I	Major-1 IDS-100 Mathematical Foundations (4T)	Minor-1 IDS-111 Problem Solving and Program Design (3T+1P)	MC-1 IDS-131 E-Commerce (3T)	AEC-1 (English-I) (2)	IDS-141 Programming in R (1T+2P)			VAC-1 (2) VAC-2 (2)	20	
II	Major-2 IDS-101 Relational Database and SQL (3T+1P) (DELETED) Major-2 IDS-102 Data Management (2T+2P)	Minor-2 IDS-112 Fundamentals of Statistics (4T)	MC-2 IDS-132 Office Software Automation (3T)	AEC-2 (MIL-I) (2)	IDS-142 Programming in Python (1T+2P)			VAC-3 (2) VAC-4 (2)	20	

III	<p>Major-3 IDS-200 Introduction to Data Science (4T)</p> <p>Major-4 IDS-201 Linear Algebra (4T)</p>	<p>Minor-3 IDS-211 Deductive and Inferential Mathematics (4T)</p> <p>OR</p> <p>IDS-212 Discrete Mathematics (4T)</p>	<p>MC-3 IDS-231 Website Design (3T)</p>	<p>AEC-3 (English-II) (2)</p>	<p>IDS-241 Basic Toolkit for Research (1T+2P)</p>			20	
IV	<p>Major-5 IDS-202 Data Modeling and Visualization (3T+1P)</p> <p>Major-6 IDS-203 Linear Programming and Optimization (4T)</p>	<p>Minor-4 VET-1 IDS-221 Rapid App Development (2T+2P) (DELETED)</p> <p>Minor-4 VET-1 IDS-222 Low code App Development (3T+1P)</p>		<p>AEC-4 (MIL-II) (2)</p>				20	

	<p>Major-7 IDS-204 Fundamentals of Computing Systems Design (4T)</p> <p>Major-8 IDS-205 LINUX Fundamentals (2P)</p>								
V	<p>Major-9 IDS-300 Introduction to Machine Learning (2T+2P)</p> <p>Major-10 IDS-301 Data Mining and Warehousing (2T+2P)</p>	<p>Minor-5 VET-2 IDS-321 Software Engineering (3T+1P)</p>				<p>IDS-361 Internship (2)</p>			20

	<p>Major-11 IDS-302 Business Data Management (4T)</p> <p>Major-12 IDS-303 Data Structures (2T)</p>								
VI	<p>Major-13 IDS-304 Cloud Computing (2T+2P)</p> <p>Major-14 IDS-305 Business Analytics and Intelligence (2T+2P)</p> <p>Major-15 IDS-306 Thematic Ideas in Data Science (4T)</p>	<p>Minor-6 VET-3 IDS-322 Big Data (3T+1P)</p> <p>OR</p> <p>IDS-323 Usability and Quality Assurance (3T+1P)</p>						20	

	<p>Major-16 IDS-307 Project (4)</p>								
VII	<p>IDS-400 Design and Analysis of Algorithms (4T)</p> <p>IDS-401 Advanced Data Management (3T+1P)</p> <p>IDS-402 Design Thinking for Data Science Applications (4T)</p> <p>IDS-403 Artificial Intelligence (3T+1P)</p>	<p>IDS-411 Programming Paradigms (3T+1P)</p> <p>OR</p> <p>IDS-412 High Performance Computing (3T+1P)</p>						20	

VIII	<p>IDS-404 Numerical Methods (4T)</p> <p>IDS-405 Building Data Science Applications (3T+1P)</p> <p>IDS-406 Predictive Analytics (3T+1P)</p> <p>IDS-407 Theory of Sampling and Design of Experiments for Data Analysis (4T)</p>	<p>IDS-413 Natural Language Processing (4T)</p> <p>OR</p> <p>IDS-414 Visual Computing (4T)</p>						20	
IX	<p>IDS-500 Research Methodology (4)</p>	<p>IDS-511 Text Mining and Sentiment Analysis (4T)</p>						20	

	IDS-501 Deep Learning (4T) RSEC Course (4) RSEC Course (4)	OR IDS-512 Image Processing (4T)							
X	RSEC Course (4)				IDS-561 Dissertation (16)			20	

List of Research Specific Elective Courses RSEC (students choose any three from a minimum of five)	
IDS-502 (Speech Processing) (4)	
IDS-503 (Recommender Systems) (4)	
IDS-504 (Financial Machine Learning) (4)	
IDS-505 (Bayesian Data Analysis) (4)	
IDS-506 (Reinforcement Learning) (4)	
IDS-507 (Time Series Analysis) (4)	
IDS-508 (Topological Data Analysis) (4)	

SEMESTER I

Major Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-100
Title of the Course : Mathematics Foundations
Number of Credits : 4
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	To introduce the students to 1. develop logical reasoning among students 2. organize all aspects of mathematics 3. understand on the fundamental concepts, assumptions and principles of mathematics	
Units	Content	No. of hours 60T
I	<p>Logic and Propositional Calculus: Propositions and Compound Statements; Basic Logical Operations; Propositions and Truth Tables; Tautologies and Contradictions; Logical Equivalence; Algebra of Propositions; Conditional and Biconditional Statements.</p> <p>Sets: Sets and their representation; The empty set; Finite and Infinite Sets; Equal Sets; Subsets; Power Set; Universal Set; Union and Intersection of sets; Venn Diagrams; Operations on Sets; Complement of a set.</p> <p>Relations and Functions: Cartesian product of sets; Relation and their types; Functions and their types; Algebra of functions; Composition of functions; Invertible functions; Binary operations.</p>	15T
II	<p>Limits: Geometric meaning of limits; Standard limits.</p> <p>Continuity: Geometric meaning of continuity; Continuous functions; Algebra of continuous functions; Examples of continuous functions; Discontinuities; Types of Discontinuities.</p> <p>Differentiability: First principle of differentiation; Algebra of differentiability namely sum/product/quotient rule; Examples; Result that every differentiable function is continuous; Derivative of the composition; Chain rule; (Statements of these results only.</p>	15T

	However, the idea of the proof, though not a part of the syllabus, is encouraged) Examples; Optimization problems.	
III	<p>Understanding Integrals integration and its relation to differentiation; antiderivatives; indefinite integrals and methods; definite integrals and interpretation as area under curve; linearity of integration; fundamental theorem of calculus</p> <p>Application of Integrals Geometric: area under curve and volumes; Physics: displacement from velocity and velocity from acceleration; Economics: consumer and producer surplus; Probability: probabilities using integrals</p>	15T
IV	<p>Complex Numbers: Algebra of complex numbers; Modulus and Complex conjugate; Argand plane and polar representation.</p> <p>Vector Algebra: Types of vectors; Addition of vectors; Multiplication of a vector by a scalar; Dot product and cross product of vectors, and their geometrical interpretation; Concept and computation of gradient, divergence, and curl of a vector field.</p>	15T
Pedagogy:	Lectures/ Case Analysis/ Assignments/ Classroom Interaction/Quiz	
References/ Reading:	<ol style="list-style-type: none"> 1. Kumar, A., Kumaresan, S., & Sarma, B. K. (2018). A Foundation Course in Mathematics. Alpha Science International. 2. Lipschutz, S., & Lipson, M. (2022). Schaum's Outline of Discrete Mathematics. 3. Mendelson, E. (2008). BEGINNING CALCULUS 3ed. McGraw-Hill. 4. Nicholson, W. K. (2020). Linear algebra with applications. 5. Spiegel, M. R., Lipschutz, S., Schiller, J. J., & Spellman, D. (2017). Schaum's outline of Complex Variables. McGraw Hill Professional. 6. Spiegel, M. R., Lipschutz, S., & Spellman, D. (2017). Vector Analysis 	
Course Outcomes:	<p>On completion of the course, students will be able to :</p> <ol style="list-style-type: none"> 1. Infer the truth of various sentences and its equivalents and outline various properties of sets. 2. Examine and Identify the types of relations and functions. 3. Make use of the strong and weak induction. 4. Solve systems of linear equations. 5. Discuss the properties of determinants. 	

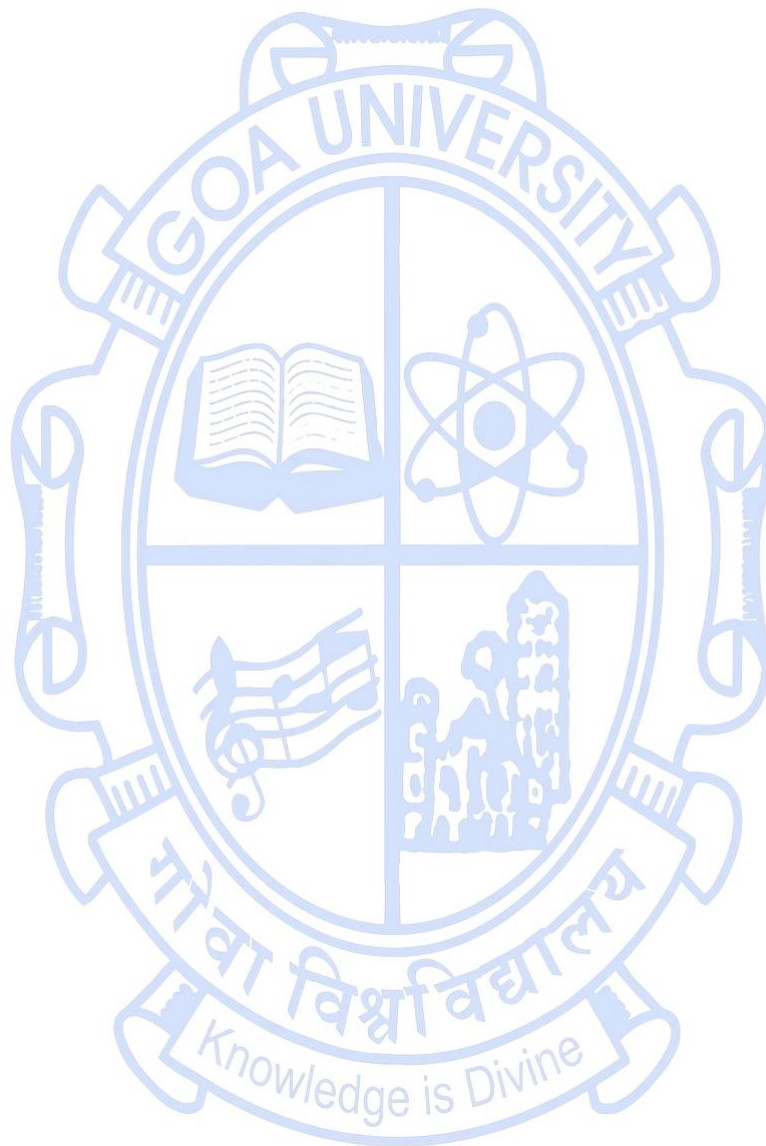
Minor Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-111
Title of the Course : Problem Solving and Program Design
Number of Credits : 4(3T+1P)
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	<ol style="list-style-type: none"> 1. To introduce the learner to various concepts and the problem solving process 2. To gain an understanding and attempt solving problems using various tools and techniques 3. To solve complex and/or large problems using composite data structures 4. To be able to design and build computer programs given a problem 	
Units	Content	No. of hours 75(45T +30P)
I	Basic Concepts <ul style="list-style-type: none"> ● Problems solving process; Solution types; Solving by computer ● Constants & Variables; Data Types & Storage; Functions; Operators; Expression & Equations ● Communicate with computer; methods to organize solution; solution planning tools; solution test & code; SDLC 	15T
II	Problem Solving Techniques <ul style="list-style-type: none"> ● Modules design; Cohesion-Coupling; variable scope, parameters & return value ● Sequential Problem Solving by charts, algorithmic instruction, flowchart symbols and pseudocode ● Decision Problem Solving by various logic types; logic conversion; introduction to decision tables ● Iterative Problem Solving; loop vs go-to; incrementing & accumulating; loop types; Recursion 	15T
III	Problem Solving with Composite Data Structures <ul style="list-style-type: none"> ● Composite data structures; heterogeneous data in structures; dynamic-sized structures ● Lists (or arrays); element & index; two-dimensional & higher; row/column-major storage; map ● Introduction to linked list, graphs, trees ● Introduction to DBMS, class/object & file processing 	15T

<p>IV</p>	<p>Practical Work</p> <p>The assignments in the practicals will reinforce problem solving, programming logic and program design concepts.</p> <p>It is recommended to use building block tools like PictoBlox, Scratch, TurtleArt, etc. and programming learning languages Kojo, Racket (DrRacket), etc.</p> <p>Following is a suggested (non-exhaustive) list of assignments:-</p> <ul style="list-style-type: none"> ● Creating simple geometric lines and shapes ● Creating complex geometric shapes ● Creating shapes for alphabets and numbers ● Creating simple drawings of everyday objects ● Moving character(s) on keyboard arrow ● Demonstrating collision between character(s) ● Creating a simple (points/score)-based game ● Write scripts to perform arithmetic operations ● Write scripts to demonstrate logical operations ● Write scripts to perform string operations ● Write scripts to calculate prime number(s), factorial(s), armstrong number(s), etc. ● Write scripts to perform geometric operations ● Write scripts to perform operation(s) on finite and infinite series' ● Write scripts demonstrating use of parentheses, and nested expressions and invocations ● Write scripts for simple games 	<p>30P</p>
<p>Pedagogy:</p>	<p>Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.</p>	
<p>References/ Readings:</p>	<ol style="list-style-type: none"> 1. Abelson, H., & Sussman, G. J. (1996). Structure and interpretation of computer programs (p. 688). The MIT Press. 2. Felleisen, M., Findler, R. B., Flatt, M., & Krishnamurthi, S. (2018). How to design programs: an introduction to programming and computing. MIT Press. 3. Kuppuswamy, S., Malliga, S., Kanimozhi Selvi, C. S., & Kousalya, K. (2019). Problem Solving and Programming. Tata McGraw Hill. 4. Maureen Sprankle, Jim Hubbard (2013). Problem Solving and Programming Concepts. Pearson Education India. 	
<p>Course Outcomes:</p>	<p>On completion of the course, students will be able to :</p> <ol style="list-style-type: none"> 1. Remember the basic concepts & terminologies of problem solving, algorithms, flowcharts, pseudo-code, and debugging. 2. Understand basic computing concepts, algorithm design, flowchart design, pseudo-code, programming constructs, and debugging. 	

- | | |
|--|---|
| | <ol style="list-style-type: none">3. Apply problem solving & programming concepts in designing solutions to simpler problems using algorithm, flowchart and pseudocode.4. Analyze a problem and apply the concepts learnt to develop & design solutions and programs |
|--|---|



Multi-disciplinary Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-131
Title of the Course : E-Commerce
Number of Credits : 3
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	1. To give fundamental understanding of e-commerce and online marketing 2. To instill ideas of Search Engine Optimization and Marketing, Applications of e-commerce and digital payments 3. To identify, define and differentiate the e-commerce models and risks of electronic commerce	
Units	Content	No. of hours 45T
I	Introduction to Electronic Commerce <ul style="list-style-type: none"> ● Basic Understanding of E-Commerce ● History of e-commerce ● Business applications of e-commerce ● E-Commerce Models (B2B, B2C, C2C, B2G) ● Applications of M-Commerce E-Commerce on Internet <ul style="list-style-type: none"> ● Websites as marketplace ● Role of website in B2C e-commerce ● Website design principles ● Alternative methods of customer communication Online Marketing <ul style="list-style-type: none"> ● Online marketing and advertising ● Push and pull approaches ● Web counters, Web advertisements ● Content marketing ● Need of Digital Marketing for a business 	15T
II	Search Engine Optimization <ul style="list-style-type: none"> ● Search Engine Optimization (SEO), Search Engine Marketing (SEM) ● Social Media Marketing (SMM), Web Analytics Applications of E-commerce <ul style="list-style-type: none"> ● Applications of e-commerce to Supply chain management ● Applications of e-commerce to Customer Relationship Management ● Product and service digitization, Remote servicing 	15T

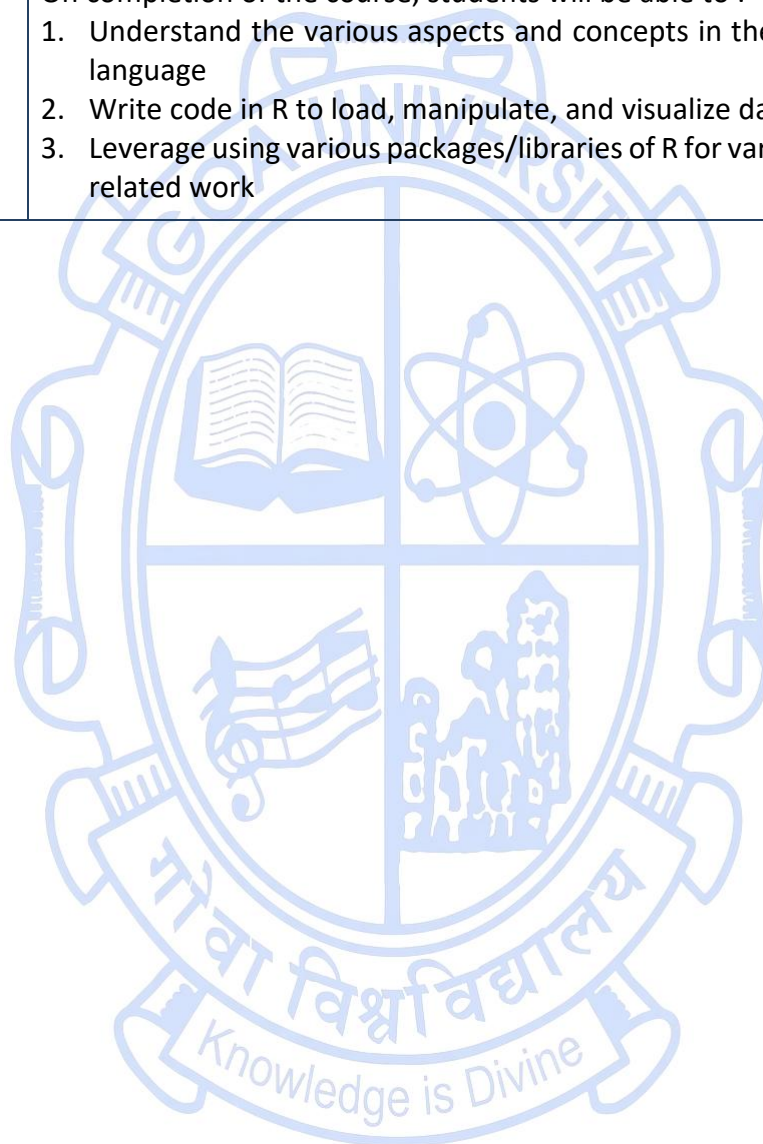
	Electronic Payment System <ul style="list-style-type: none"> Types of payment systems, credit cards, debit cards, mobile, etc., Electronic Fund Transfer (EFT) Operational credit and legal risk of e-payment, Risk management options for e-payment systems 	
III	Business to Consumer E-Commerce <ul style="list-style-type: none"> Cataloging Order planning and order generation, Cost estimation and pricing Order receipt and accounting, Order selection and prioritization Order scheduling, Order fulfilling, Order delivery Order billing, Post sales service Business to Business E-Commerce <ul style="list-style-type: none"> Need and Models of B2B e-commerce Using public and private computer networks for B2B trading EDI and paperless trading, Characteristic features of EDI service arrangement, EDI architecture and standards Security Issues in E-Commerce <ul style="list-style-type: none"> Risks of e-commerce Types and sources of threats; Security tools Risk management approaches 	15T
Pedagogy:	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading:	<ol style="list-style-type: none"> Kalakota, Ravi, Andrew Whinston (2015). Frontiers of Electronic Commerce. Pearson Education P. T. Joseph (2015). E-Commerce: An Indian Perspective Paperback. PHI Learning. V. Rajaraman (2015). Essentials of E-Commerce Technology. PHI Learning. 	
Course Outcomes:	On completion of the course, students will be able to: <ol style="list-style-type: none"> Understand the foundation of e-commerce, e-commerce websites and Online Marketing and Security Issues Explain the importance of Search Engine Optimization, Applications of E-commerce and Electronic Payment Systems Have an in-depth understanding of B2B and B2C e-commerce models 	

Skill Enhancement Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-141
Title of the Course : Programming in R
Number of Credits : 3 (1T+2P)
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	1. To understand critical programming language concepts 2. To make use of R loop functions and debugging tools 3. To configure and use various libraries & packages with R	
Units	Content	No. of hours 75(15T +60P)
I	The R programming language <ul style="list-style-type: none"> ● R language; variables, constants & operators; data types, objects & in-built library calls ● Decision & iterative structures; functions ● Vectors, lists, matrices; arrays; data frames, factors ● Visualization using various charts/plots ● Statistical calculations ● Packages & libraries with R 	15T
II	Practical Work (Basic Programming in R) <ul style="list-style-type: none"> ● Assignments to practice arithmetic operations and expressions ● Assignments to practice working with an infinite series ● Assignments to practice working with arrays, matrix, vectors, list, factors ● Assignments to practice with dataframe ● Assignments to work with file processing ● Assignments to practice with various statistical calculations 	30P
III	Practical Work (Visualization, Packages in R) <ul style="list-style-type: none"> ● Assignments to practice with various charts / graphs ● Assignments to use package(s) related to loading, manipulating & modeling data ● Assignments to use package(s) related to visualizing & reporting data ● Assignments to use package(s) related to work with spatial, time-series and financial data 	30P
Pedagogy :	Lectures, Tutorials. Assignments, Seminars, Presentations or a	

	combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.
References/ Reading:	<ol style="list-style-type: none"> 1. Crawley, M. J. (2012). The R book. John Wiley & Sons. 2. Wickham, H., Çetinkaya-Rundel, M., & Golemund, G. (2023). R for data science. " O'Reilly Media, Inc.". 3. https://cran.r-project.org/manuals.html [Accessed: Jan 20, 2024]
Course Outcomes:	<p>On completion of the course, students will be able to :</p> <ol style="list-style-type: none"> 1. Understand the various aspects and concepts in the R programming language 2. Write code in R to load, manipulate, and visualize data 3. Leverage using various packages/libraries of R for various data science related work



SEMESTER II

Major Courses

Name of the Programme : Master of Science (Integrated) in Data Science

Course Code : IDS-102

Title of the Course : Data Management

Number of Credits : 2T+2P

Effective from AY : 2025-26

Pre-requisites for the Course:	NIL	
Course Objective :	<ul style="list-style-type: none">• Enables the learner to understand the different issues involved in the design and implementation of a database system• Provides both theoretical knowledge and practical skills required in the creation and management of a Relational DataBase Management System.	
Units	Content	No. of hours 90 (30T+60P)
I	Foundational Concepts <ul style="list-style-type: none">• Understanding Data; Database (DB) & DB Users; Characteristics; Structured and Unstructured Databases• DB Architectures; Data Models; Schemas & Instances• Data Modeling using the ER approach• Relational Data Model & Introduction to Relational Algebra• Relational DB Design; Functional Dependencies;• Introduction to Normalization & Basic Normal Forms• Basic structure of Database Queries• SQL - DDL• Set Operations• Null Values and Aggregate Functions	15T
II	Advanced SQL <ul style="list-style-type: none">• Nested Subqueries• Joins• Correlated Subqueries• Database Modification• Introduction to Advanced SQL	15T
III	Practical Work Foundational Concepts <ul style="list-style-type: none">• Getting Started<ul style="list-style-type: none">○ Installation of DBMS Software(s)	30P

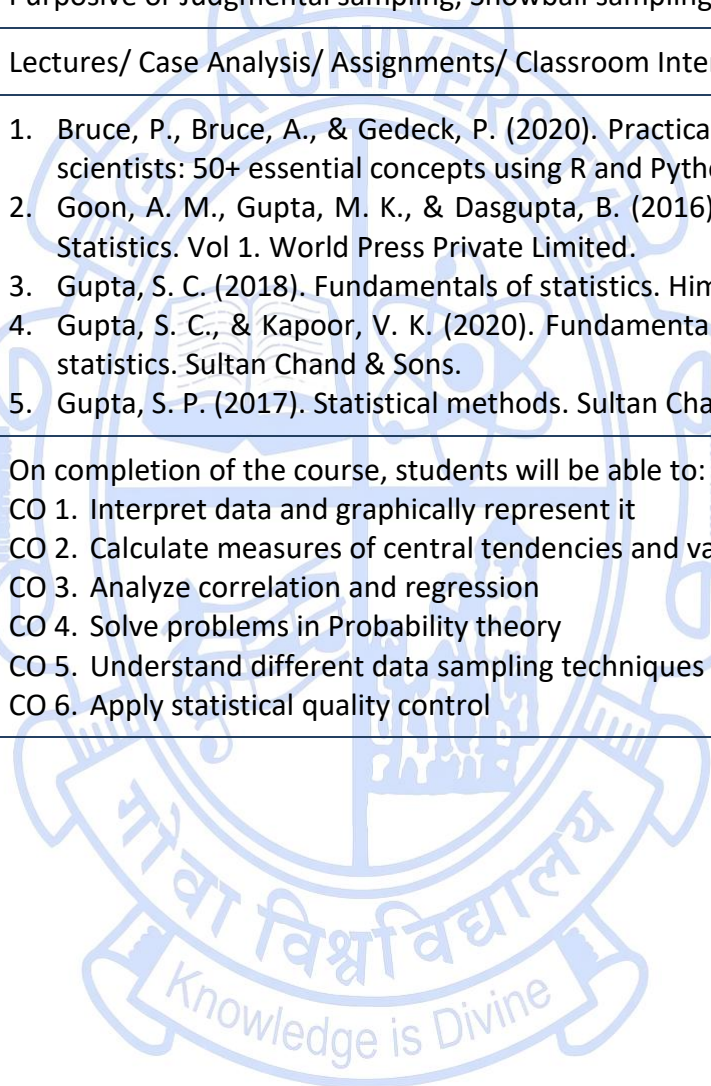
	<ul style="list-style-type: none"> ○ Creation/modification of database tables using DDL statements and GUI tools of the DBMS software ○ Populating tables, Adding Primary Keys, Candidate Keys, Foreign Keys and other Integrity Constraints to relations ○ Updating and Deleting Rows of tables ○ Altering schema of existing tables ○ Writing simple SQL queries ● Querying Data <ul style="list-style-type: none"> ○ Querying the data dictionary ○ Various Operators ○ Using alias to control column headings ○ Using compound clauses and wildcards ○ Sorting data ○ Arithmetic Operators and Built-in Functions 	
IV	Practical Work Advanced SQL <ul style="list-style-type: none"> ● Writing Complex Queries <ul style="list-style-type: none"> ○ Joins ○ Aggregate Functions ○ Group by and Having Clause ○ Subqueries – single row and multiple rows ○ Correlated Subqueries 	30P
Pedagogy :	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading :	<ol style="list-style-type: none"> 1. Connolly, T. M., & Begg, C. E. (2005). Database systems: a practical approach to design, implementation, and management. Pearson Education. 2. Elmasri, R., Navathe, S. B., Elmasri, R., & Navathe, S. B. (2015). Fundamentals of Database Systems. (7th Edition). Pearson Publisher 3. Garcia-Molina, H. (2008). Database systems: the complete book. Pearson Education India. 4. Ramakrishnan, R., Gehrke, J., & Gehrke, J. (2003). Database management systems (Vol. 3). New York: McGraw-Hill. 5. Silberschatz, A., Korth, H. F., & Sudarshan, S. (2011). Database system concepts. (7th Edition). McGraw-Hill. 	
Course Outcomes:	On completion of the course, students will be able to : CO1: Understand and evaluate the role of a DBMS in Organizations. CO2: Recognize a good design and understand the relational database design principles. CO3: Use logical design methods used in the design of DB applications. CO4: Understand the basics of SQL and construct queries using SQL.	

Minor Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-112
Title of the Course : Fundamentals of Statistics
Number of Credits : 4
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL		
Course Objectives:	To introduce the students to 1. organize, summarize and analyze data 2. draw appropriate conclusions from data 3. apply statistical tools and techniques to real-world problems		
Units	Content	No. of hours	
I	<p>Introductory concepts: Definition and scope of Statistics; Concept of population and sample.</p> <p>Types of data: Quantitative; Qualitative; Attributes; Variates.</p> <p>Tabulation of data: Class intervals; Frequency tables.</p> <p>Presentation of data: Diagrams and graphs: Bar diagrams and their types; Pie charts; Frequency polygon; Histogram; Ogives.</p> <p>Consistency and independence of data with special reference to attributes.</p> <p>Scales of measurement: Nominal, Ordinal, Interval, Ratio.</p> <p>Measures of Central Tendency: Mathematical and Positional – Mean, Median, Mode, Quartiles, Percentiles.</p> <p>Measures of Dispersion: Range, Quartile deviation, Standard deviation, Coefficient of variation.</p>	60	
II	<p>Bivariate data: Definition; Scatter diagram.</p> <p>Correlation and Regression: Simple, Partial and Multiple Correlation (3 variables only); Rank correlation; Simple linear regression.</p>	15T	
III	<p>Probability: Introduction; Random experiments; Sample space; Events and algebra of events; Definitions of Probability – Classical, Statistical, and Axiomatic; Conditional Probability; Addition and Multiplication theorem of probability; Independent events; Theorem of Total probability; Bayes' theorem and its applications.</p>	15T	

IV	<p>Statistical Quality Control: Introduction; Causes of variation in quality; Objective, advantages, and techniques of SQC.</p> <p>Attribute data: P chart, U chart, C chart.</p> <p>Numerical data: X bar chart, R bar chart, S bar chart.</p> <p>Sampling techniques: Various methods of data collection; Census survey and sample survey.</p> <p>Sampling Methods: Simple random sampling; Systematic sampling; Stratified sampling; Clustered sampling.</p> <p>Non – probability Sampling Methods: Convenience sampling; Consecutive sampling; Quota sampling; Purposive or Judgmental sampling; Snowball sampling.</p>	15T
Pedagogy:	Lectures/ Case Analysis/ Assignments/ Classroom Interaction/ Quiz	
References/ Reading:	<ol style="list-style-type: none"> 1. Bruce, P., Bruce, A., & Gedeck, P. (2020). Practical statistics for data scientists: 50+ essential concepts using R and Python. O'Reilly Media. 2. Goon, A. M., Gupta, M. K., & Dasgupta, B. (2016). Fundamentals of Statistics. Vol 1. World Press Private Limited. 3. Gupta, S. C. (2018). Fundamentals of statistics. Himalaya Pub. 4. Gupta, S. C., & Kapoor, V. K. (2020). Fundamentals of mathematical statistics. Sultan Chand & Sons. 5. Gupta, S. P. (2017). Statistical methods. Sultan Chand & Sons. 	
Course Outcomes:	<p>On completion of the course, students will be able to:</p> <p>CO 1. Interpret data and graphically represent it</p> <p>CO 2. Calculate measures of central tendencies and variations</p> <p>CO 3. Analyze correlation and regression</p> <p>CO 4. Solve problems in Probability theory</p> <p>CO 5. Understand different data sampling techniques</p> <p>CO 6. Apply statistical quality control</p>	



Multi-disciplinary Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-132
Title of the Course : Office Software Automation
Number of Credits : 3T
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	1. To understand the basics of office automation software and its applications. 2. To develop proficiency in using word processing, spreadsheet, and presentation software. 3. To diagnose and troubleshoot common PC issues and optimize the performance of a PC.	
Units	Content	No. of hours 45
I	Basic Concepts <ul style="list-style-type: none"> Understanding office automation software and its applications; Types of office automation software Office Suites like (Libre Office, Microsoft Office Suite, Google Workspace) Spreadsheets <ul style="list-style-type: none"> Introduction to spreadsheet software (LibreOffice Calc, MS Excel, Google Sheet) Creating and formatting spreadsheets Working with formulas and functions Charts and graphs; Collaboration tools 	15T
II	Word Processing <ul style="list-style-type: none"> Introduction to word processing software (LibreOffice Writer, MS Word, Google Doc) Creating and formatting documents Working with templates Mail merge and labels; Collaboration tools Presentation Software <ul style="list-style-type: none"> Introduction to presentation software (LibreOffice Impress, MS Powerpoint, Google Slide) Creating and formatting presentations Working with images, videos, and animations Collaboration tools 	15T
III	Internet and Email <ul style="list-style-type: none"> Introduction to the Internet; Web Browsers; Searching 	15T

	<p>Engines; Configuring web browsers</p> <ul style="list-style-type: none"> ● Introduction to Email; Setting up and configuring email accounts; Composing and sending emails; Managing Email Accounts <p>PC Troubleshooting</p> <ul style="list-style-type: none"> ● Hardware Troubleshooting: Basic hardware components of a PC, Common hardware issues and their solutions, Maintenance and optimization of hardware ● Software Troubleshooting: Common software issues and their solutions, Malware and virus removal, System recovery and backups ● Network Troubleshooting 	
Pedagogy:	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading:	<ol style="list-style-type: none"> 1. A+ Guide to IT Technical Support (MindTap Course List) by Jean Andrews 2. Discovering Computers 2022: Digital Technology, Data, and Devices by Misty E. Vermaat 3. Meyers, M. (2017). CompTIA A+ Certification All-in-One Exam Guide, Ninth Edition (Exams 220-901 & 220-902). McGraw Hill Professional. 4. Russel, C., & Hoque, M. R. (2018). Google Workspace for Dummies. John Wiley & Sons. 5. Shelly, G. B., & Vermaat, M. E. (2017). Microsoft Office 365 & Office 2016: Introductory. Cengage Learning. 	
Course Outcomes:	<p>On completion of the course, students will be able to:</p> <p>CO 1. To create and format documents, create and format tables and mail merge</p> <p>CO 2. Understand the use and various functions of spreadsheets</p> <p>CO 3. Apply the knowledge of tools to create effective presentations</p> <p>CO 4. Understand PC assembling and troubleshooting</p>	

Skill Enhancement Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-142
Title of the Course : Programming in Python
Number of Credits : 3(1T+2P)
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL		
Course Objectives:	1. To understand critical programming language concepts 2. To make use of python loop functions and debugging tools 3. To write program using class and objects 4. To configure and use various libraries & packages with python		
Units	Content	No. of hours 75(15T +60P)	
I	The Python programming language <ul style="list-style-type: none"> Python language; variables, constants & operators; data types, objects & in-built library calls Selection and Iterative Constructs: Writing conditions, IF-ELSE constructs Conditional operators, SWITCH ,WHILE and FOR loops, Use of BREAK and CONTINUE statements. Nested Loops Advance Data types: Lists, Tuples, Set, Dictionaries, Strings, Unicode, formatting strings, docString. Searching and sorting algorithms without using library functions. Modular Programming: Importance of User Defined Functions, Hierarchy charts, fan-in/out, cohesion and coupling and loosely coupled modules. Fan-in – Fan-out concepts. User Defined Functions: Local and Global Variables, Scoping Rules, Parameters & arguments. Function with variable arguments. Modules, packages, scope. Recursion & Recursive Functions. Recursive v/s Iterative Functions. Custom Data Types and File Management: Object of a Class and basic concept of classes & OOP, Files, Exceptions in file handling. Introduction to Packages: Python packages for plotting, mathematical computation & linear regression. 	15T	
II	Practical Work (Basic Programming in Python) <ul style="list-style-type: none"> Introduction to UNIX environment- Introduction to Fedora/Ubuntu, Basic directory and file handling commands, Editor (vi editor), man pages, installation of 	30P	

	<p>Python and Jupyter notebook.</p> <ul style="list-style-type: none"> ● Assignments to practice decision control, branch and loop control structure ● Assignment to practice List, Set, Tuple, Dictionary & Strings ● Assignments to practice functions & Recursion ● Assignments to practice user-defined data types & file handling ● Assignments to work with reading and writing data to file with exception handling 	
III	<p>Practical Work (Packages in Python)</p> <ul style="list-style-type: none"> ● Assignment to create custom datatype using classes and objects ● Assignments to use package(s) related to loading, manipulating data using pandas ● Assignment to use Linear regression ● Assignments to practice with various charts / graphs/3-D charts/boxplot 	30P
Pedagogy:	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading:	<ol style="list-style-type: none"> 1. Guttag, J. V. (2016). Introduction to computation and programming using Python: With application to understanding data. MIT press. 2. Kumar, N., & Taneja, S. (2018). Python Programming: A Modular Approach. 3. Sprankle, M., & Hubbard, J. (2012). Problem solving and programming concepts. Prentice Hall Press. 4. https://scikit-learn.org/stable/index.html [Accessed: Jan 20, 2024] 	
Course Outcomes:	<p>On completion of the course, students will be able to:</p> <p>CO 1. Analyze a given problem and develop a Python program to solve it.</p> <p>CO 2. Identify test cases for a given problem.</p> <p>CO 3. Understand, test, trace programs written in Python language.</p> <p>CO 4. Working with python Standard Libraries</p>	

SEMESTER III

Major Courses

Name of the Programme : Master of Science (Integrated) in Data Science

Course Code : IDS-200

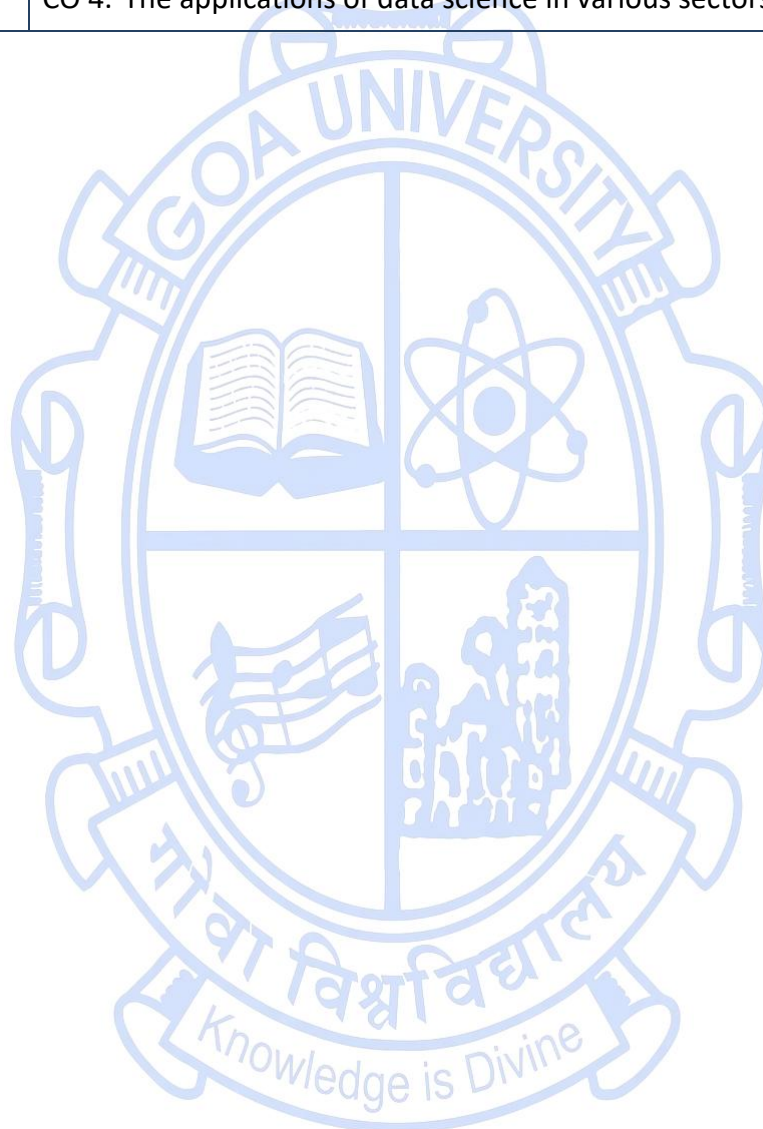
Title of the Course : Introduction to Data Science

Number of Credits : 4

Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	To gain a thorough understanding of the data science fundamentals.	
Units	Content	No. of hours 60
I	Introduction - meaning, definition, significance, Role of data scientist, Data Science Process Life Cycle, Challenges in data science process, applications, the significance of domain knowledge in data science, the significance of mathematics in data science	15T
II	Data - information vs data, structured and unstructured data, sources of data, data collection methods. Data preprocessing - cleaning, integration, transformation, reduction, discretization, feature selection and extraction, scaling, normalizing, dimensionality reduction.	15T
III	Data analysis - descriptive analysis, diagnostic analysis, predictive analysis, exploratory analysis. Model building basics - linear regression, classification, trend analysis. Model Evaluation - accuracy, precision, recall, F1-score.	15T
IV	Ethics, bias, and privacy in data science. Information security. Applying data science to real-world problems. Case studies from industry - Business, Healthcare, Finance, Education, Tourism, etc.	15T
Pedagogy:	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	

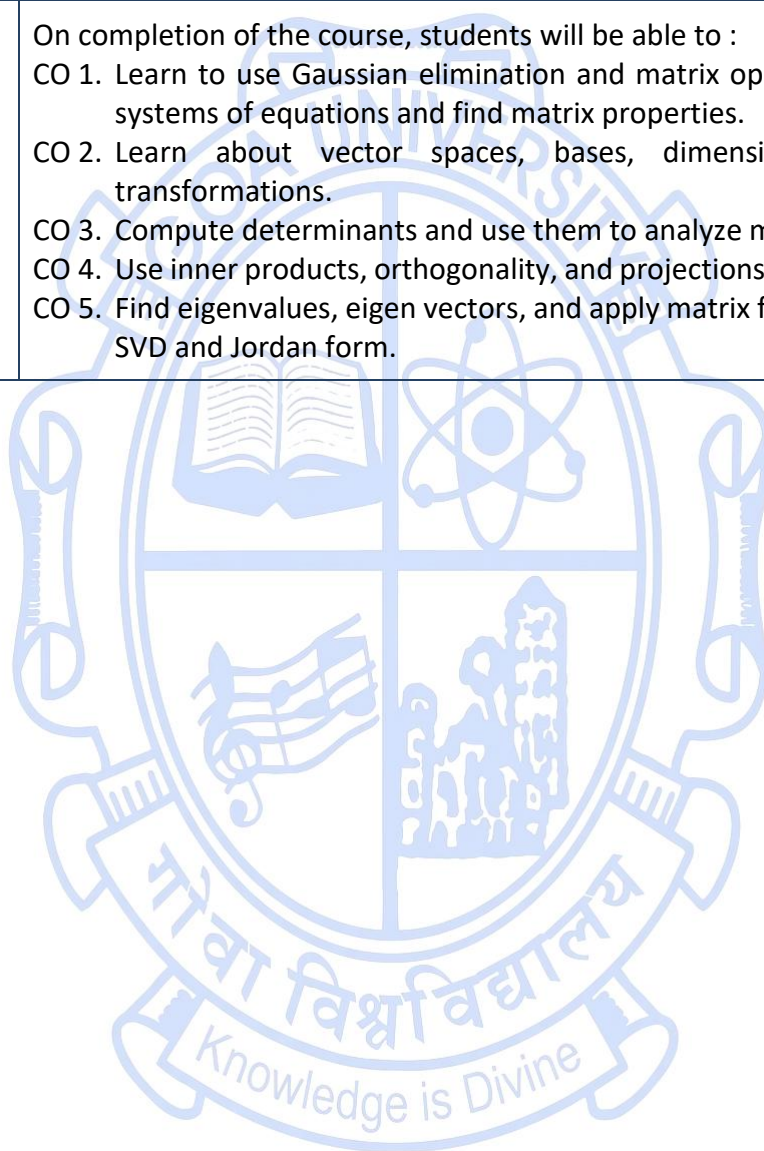
References/ Reading:	<ol style="list-style-type: none"> 1. Kelleher, J. D., & Tierney, B. (2018). Data science. MIT press. 2. Pierson, L. (2021). Data science for dummies. John Wiley & Sons. 3. Blum, A., Hopcroft, J., & Kannan, R. (2020). Foundations of data science. Cambridge University Press.
Course Outcomes:	<p>On completion of the course, students will be able to understand:</p> <p>CO 1. The role of data science in other domains.</p> <p>CO 2. The data science process/pipeline for a given problem.</p> <p>CO 3. Data analytics techniques to collect and process data.</p> <p>CO 4. The applications of data science in various sectors.</p>



Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-201
Title of the Course : Linear Algebra
Number of Credits : 4
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	1. Understand fundamental concepts of linear algebra, including matrices, determinants, vector spaces, and transformations. 2. Develop problem-solving skills using methods like Gaussian elimination, eigenvalues, and singular value decomposition. 3. Apply linear algebra to real-world problems in computing, data science, and engineering.	
Units	Content	No. of hours 60
I	Linear Systems: Systems of linear equations, elementary row operations, rank of a matrix, Gaussian elimination, elementary matrices, inverse of a matrix. Determinants: Determinant as area and volume, determinant using permutations, properties of determinants, Singular and Non-singular matrices, Rank of a matrix	15T
II	Vector Spaces: Vector Spaces over fields, subspaces, linear independence, bases, dimension. Linear Transformations: Linear transformations, algebra of linear transformations, Rank Nullity Theorem and applications, isomorphism, change of bases, transpose of a linear transformation.	15T
III	Inner Product Spaces: Inner products, Gram-Schmidt orthogonalization, orthogonal projections and best approximation, linear functionals and adjoint operator, bilinear maps, quadratic forms, symmetric, hermitian, unitary and normal operators.	15T
IV	Eigenvalues and Eigenvectors: Eigenvalues and eigenvectors, characteristic polynomial, minimal polynomials, Cayley-Hamilton Theorem, triangulation and	15T

	diagonalization, Finite dimensional spectral theorem for normal operators, Singular Value Decomposition, Jordan canonical form.
Pedagogy:	Lectures/ Case Analysis/ Assignments/ Classroom Interaction/Quiz
References/ Reading:	<ol style="list-style-type: none"> 1. Lang, S. (1987). Linear algebra. Springer. 2. Hoffman, K., & Kunze, R. (1978). Linear algebra (2nd ed.). Prentice Hall India Learning Private Limited. 3. Axler, S. (1997). Linear algebra done right (2nd ed.). Springer.
Course Outcomes:	<p>On completion of the course, students will be able to :</p> <p>CO 1. Learn to use Gaussian elimination and matrix operations to solve systems of equations and find matrix properties.</p> <p>CO 2. Learn about vector spaces, bases, dimension, and linear transformations.</p> <p>CO 3. Compute determinants and use them to analyze matrices.</p> <p>CO 4. Use inner products, orthogonality, and projections in vector spaces.</p> <p>CO 5. Find eigenvalues, eigen vectors, and apply matrix factorizations like SVD and Jordan form.</p>



Minor Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-211
Title of the Course : Deductive and Inferential Mathematics
Number of Credits : 4
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	1. To use logical reasoning to solve problems and prove statements. 2. To earn key mathematical concepts like logical equivalence and normal forms. 3. To apply probability and statistics to analyze data and make predictions.	
Units	Content	No. of hours 60
I	Mathematical Logic & Proofs An open sentence, a closed sentence, Definition of proposition or a Statement. Strong emphasis on the Distinction between Inclusive OR and Exclusive OR. Truth tables. Compound Proposition. Algorithms, Truth tables and Tautologies Equivalent statements (\equiv). Examples and important logical results. De Morgan Laws for negation. Conditional and Biconditionals, Arguments and Proofs	15T
II	Well-formed-formulae Equivalence of formulae. Various laws governing the well-formed formulae. Duality law. Normal Form. Disjunctive normal form, conjunctive normal form, Principal disjunctive normal form, Principal conjunctive normal form. Propositional Calculus. Predicate Calculus. Predicate Formula. Equivalence of Predicate Formulae. Inference Theory	15T
III	Introduction to Inference Theory Introduction to Probability Theory using Kolmogorov Technique: Definition of an experiment. Outcomes of an experiment. Outcomes which are not decomposable. Sample space as the set of all non-decomposable outcomes of an experiment. Event as any subset of the sample space of an experiment under consideration. Idea of variations. Standard deviation as the root mean square deviation with respect to the mean. Mathematical Expectation and Expected Values.	15T

	Random Variables: Idea of Distribution of a Function. Some standard Distributions such as Binomial., Normal, Poisson and Exponential. Their standard properties with the stress on Normal Distribution. Use of Normal Distribution Table to solve problems.	
IV	Logic Programming Propositional Logic vs. Predicate Logic; Fundamentals of Logical Knowledge Representation (Facts, Rules, and Queries); Unification in logic programming (predicates, variables, constants, logical operators); Recursive rule definitions; solving logical inference using resolution; solving constraint satisfaction problems	15T
Pedagogy:	Lectures/ Practice Problems/ Assignments/ Classroom Interaction/ Quiz	
References/ Reading:	<ol style="list-style-type: none"> 1. Tarski, A. (1994). Introduction to Logic and to the Methodology of the Deductive Sciences (Vol. 24). Oxford university press. 2. Sahu, P. K., Pal, S. R., & Das, A. K. (2015). Estimation and inferential statistics (pp. 1-317). New Delhi: Springer. 3. Ben-Ari, M. (2012). Mathematical logic for computer science. Springer Science & Business Media. 4. Tremblay, J. P., & Manohar, R. (2017). Discrete mathematical structures with applications to computer science. McGraw-Hill, Inc. 	
Course Outcomes:	<p>On completion of the course, students will be able to:</p> <p>CO 1. Explore, conjecture, and reason logically to model and arrive at a solution to a given problem.</p> <p>CO 2. Apply different mathematical methods to solve problems effectively.</p> <p>CO 3. Learn decision-making in the presence of uncertainty and will learn to quantify the uncertainty in estimation /the decision</p>	

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-212
Title of the Course : Discrete Mathematics
Number of Credits : 4
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	1. To understand fundamental concepts of, sets, relations, and functions for mathematical reasoning. 2. To develop problem-solving skills using combinatorics, graph theory, and recursion. 3. To apply discrete mathematics to algorithms, data structures, and computational problems in computer science.	
Units	Content	No. of hours 60
I	Mathematical Logic: Introduction, Statements and Notation, Connectives, Well-formed formulas, Tautology, Duality law, Equivalence, Implication, Normal Forms, Functionally complete set of connectives, Inference Theory of Statement Calculus, Predicate Calculus, Inference theory of Predicate Calculus. Set Theory: Basic Concepts of Set Theory, Relations and Ordering, The Principle of Inclusion-Exclusion, Pigeonhole principle and its application, Functions, Composition of functions, Inverse Functions, Recursive Functions, Lattices and its properties.	15T
II	Algebraic Structures: Examples and General Properties, Semi-groups and Monoids, Groups, Sub-groups, Homomorphism, Isomorphism, Cosets, Lagrange's Theorem, Group Actions, Cyclic Groups, Abelian Groups, Permutation Groups, Symmetry Groups, Properties of groups and their applications, Normal Subgroups, Quotient Groups, Applications in cryptography and coding theory.	15T
III	Elementary Combinatorics and Recurrence Relations Basics of Counting, Combinations and Permutations, Enumeration of Combinations and Permutations, Enumerating Combinations and Permutations with Repetitions, Enumerating Permutations with Constrained Repetitions, Binomial Coefficients, The Binomial and Multinomial Theorems, Advanced Counting Techniques, Pigeonhole Principle, Inclusion-Exclusion Principle, Stirling Numbers, Recurrence Relations, Generating	15T

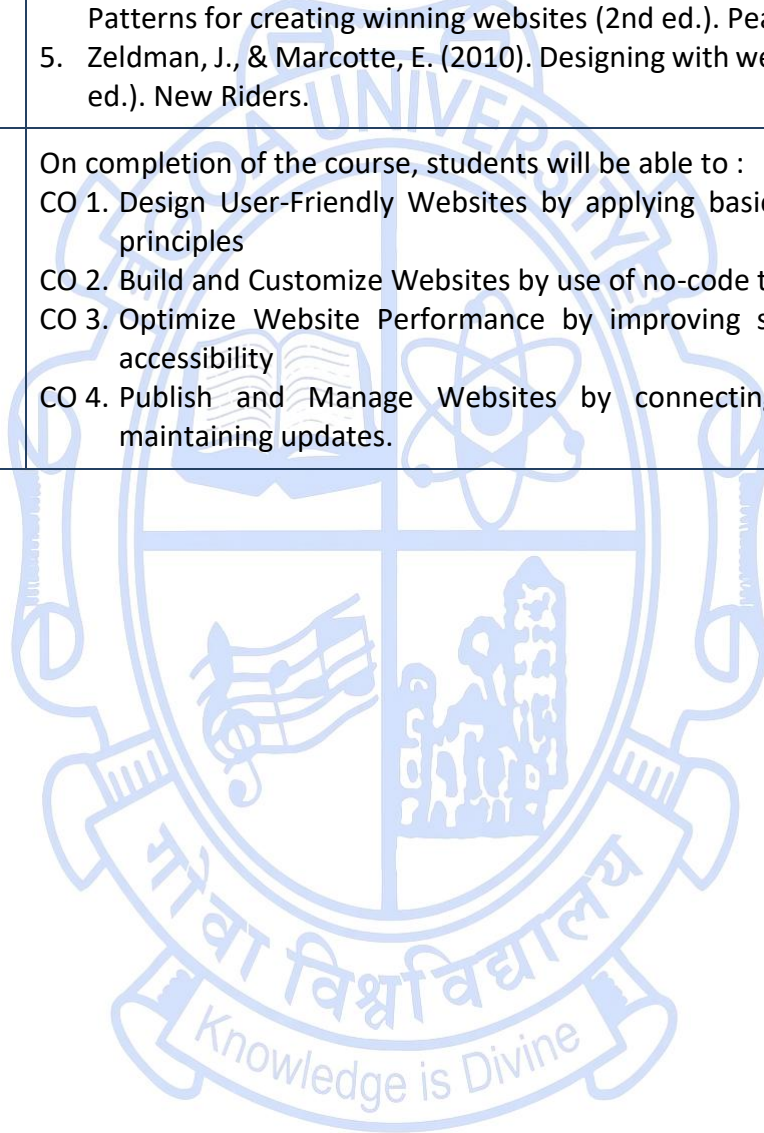
	Functions of Sequences, Calculating Coefficients of Generating Functions, Solving Recurrence Relations by Substitution and Generating Functions, The Method of Characteristic Roots, Solutions of Inhomogeneous Recurrence Relations.	
IV	<p>Graph Theory</p> <p>Basic Concepts of Graphs, Types of Graphs, Isomorphism and Sub-graphs, Graph Representations (Adjacency Matrix, Adjacency List), Trees and their Properties, Spanning Trees, Directed Trees, Binary Trees, Planar Graphs, Euler's Formula, Multigraphs and Euler Circuits, Hamiltonian Graphs, Chromatic Numbers, The Four Color Problem, Graph Traversals (DFS, BFS), Network Flow, Bipartite Graphs, Graph Coloring, Applications of Graphs in Networks and Scheduling.</p>	15T
Pedagogy:	Lectures/Practice Problems/Assignments/ Classroom Interaction/ Quiz	
References/ Reading:	<ol style="list-style-type: none"> 1. Tremblay, J. P., & Manohar, R. (2017). Discrete mathematical structures with applications to computer science. McGraw-Hill, Inc. 2. Kolman, B., Busby, R. C., & Ross, S. C. (2011). <i>Discrete mathematical structures</i>. PHI Learning. 3. Malik, D. S., & Sen, M. K. (2004). <i>Discrete mathematical structures: Theory & applications</i>. Thomson India Edition. 	
Course Outcomes:	<p>On completion of the course, students will be able to:</p> <p>CO 1. Identify and work with finite, infinite, and uncountable sets, apply mathematical induction, and use inclusion-exclusion principles.</p> <p>CO 2. Solve problems involving permutations, combinations, discrete, and conditional probability.</p> <p>CO 3. Apply properties of binary relations and functions, including in databases and job scheduling.</p> <p>CO 4. Work with graphs, Eulerian/Hamiltonian paths, spanning trees, and planar graphs.</p> <p>CO 5. Algebraic Structures and Boolean Algebra: Understand group theory, rings, fields, and manipulate Boolean functions and propositional calculus.</p>	

Multi-disciplinary Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-231
Title of the Course : Website Design
Number of Credits : 3
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	<ol style="list-style-type: none"> 1. Understand website basics, design principles, and no-code tools. 2. Learn how to structure content, use media, and improve usability. 3. Optimize, maintain, and update websites for better user experience and visibility 	
Units	Content	No. of hours 45
I	Introduction to Website Design <ol style="list-style-type: none"> 1. Website concept; types of websites; components of website and webpage; UX & UI; accessibility and mobile responsiveness 2. Principles of good design; wireframing and prototyping 3. No-code website builders; choosing/customizing template; adding pages 	15T
II	Building and Structuring a Website <ol style="list-style-type: none"> 1. Structuring a website with navigation menu, content writing and placement, organize sections 2. Adding basic interactive features like embedded media, contact forms, social media integrations 3. Publish/manage website by understanding domain/hosting, basics of SEO, and launching personal website 	15T
III	Enhancing and Maintaining Websites <ol style="list-style-type: none"> 1. Website optimization by improving website loading speed, and image/media optimization without quality loss 2. Security basics: Backups, SSL certificates, and privacy policies 3. Advanced Features like creating forms, pop-ups, and live chat, embedding Google Analytics for visitor insights, simple automation like email signup, appointment booking 4. Website maintenance by regular website updates, monitoring visitor engagement, digital marketing 	15T
Pedagogy:	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group	

	learning.
References/ Reading:	<ol style="list-style-type: none"> 1. Krug, S. (2014). Don't make me think, revisited: A common sense approach to web usability (3rd ed.). New Riders. 2. McFarland, D. S. (2021). CSS: The missing manual (5th ed.). O'Reilly Media. 3. Robbins, J. N. (2018). Learning web design: A beginner's guide to HTML, CSS, JavaScript, and web graphics (5th ed.). O'Reilly Media. 4. Van Duyne, D. K., Landay, J. A., & Hong, J. I. (2006). The design of sites: Patterns for creating winning websites (2nd ed.). Pearson Education. 5. Zeldman, J., & Marcotte, E. (2010). Designing with web standards (3rd ed.). New Riders.
Course Outcomes:	<p>On completion of the course, students will be able to :</p> <p>CO 1. Design User-Friendly Websites by applying basic design and UX principles</p> <p>CO 2. Build and Customize Websites by use of no-code tools</p> <p>CO 3. Optimize Website Performance by improving speed, SEO, and accessibility</p> <p>CO 4. Publish and Manage Websites by connecting domains and maintaining updates.</p>



Skill Enhancement Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-241
Title of the Course : Basic Toolkit for Research
Number of Credits : 3(1T+2P)
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	1. To have basic knowledge of various software tools and platforms for data handling. 2. To gain hands-on experience with various software tools for research.	
Units	Content	No. of hours 75 (15T+60P)
I	<p>Markdown Document structure; basic text formatting; paragraphs; headings; lists; links and images; code blocks; escape characters; HTML elements; converting markdown to HTML web pages.</p> <p>Source Version Control Version Control; introduction to SVN and Git; Git repositories; Git cloning, forks/branches; Git stash; Git pull requests; Git merge conflicts; Git pages.</p> <p>Scientific Documents TEX; LaTeX; documents, lists, enumeration, figures/tables; mathematics; symbols; referencing</p>	15T
II	<p>Practical Work Markdown - Setting up portfolio, creating code chunk, formatting, adding links, images, tables, tables, mathematical formulae.</p> <p>Git basics - Setting up a repository, managing branches, remote repositories, analysing git history, push code to GitHub, cherry-pick a range of commits, using git to track changes</p>	30P
III	<p>Practical Work Latex - Document structure, lists and enumeration, inserting figures, inserting tables, including mathematical symbols and</p>	30P

	<p>equations, cross-referencing, bibliography and citation, switching templates.</p> <p>IDEs (RStudio/Spyder) - Setting up of IDE, features, exploring the interface, installing and loading packages, using Markdown for documentation.</p> <p>Python Notebooks/ Google Colab: Installation, cell types Code, Markdown, Raw NBConvert, code and syntax highlighting, adding forms in colab, exporting to different formats, extensions, publishing and sharing the notebook using free GPU.</p>	
Pedagogy:	Lectures, Practical sessions, Assignments, Seminars, Presentations, Mini-projects or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading:	<ol style="list-style-type: none"> 1. Cone, M. (2020). Markdown guide. Independently Published. 2. Scott, C., & Ben, S. (2016). Pro Git. 3. Oetiker, T., Partl, H., Hyna, I., & Schlegl, E. (2014). The not so short introduction to LaTeX. Typeset LATEX 2ε. 4. Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). R for data science. " O'Reilly Media, Inc." 	
Course Outcomes:	<p>On completion of the course, students will be able to :</p> <p>CO 1. Use Markdown for effective documentation.</p> <p>CO 2. Use LaTeX for research documentation</p> <p>CO 3. Collaborate on data analysis projects using version control tool.</p> <p>CO 4. Make use of IDEs and python notebooks for programming and research.</p>	

SEMESTER IV

Major Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-202
Title of the Course : Data Modeling and Visualization
Number of Credits : 4 (3T+1P)
Effective from AY : 2024-25

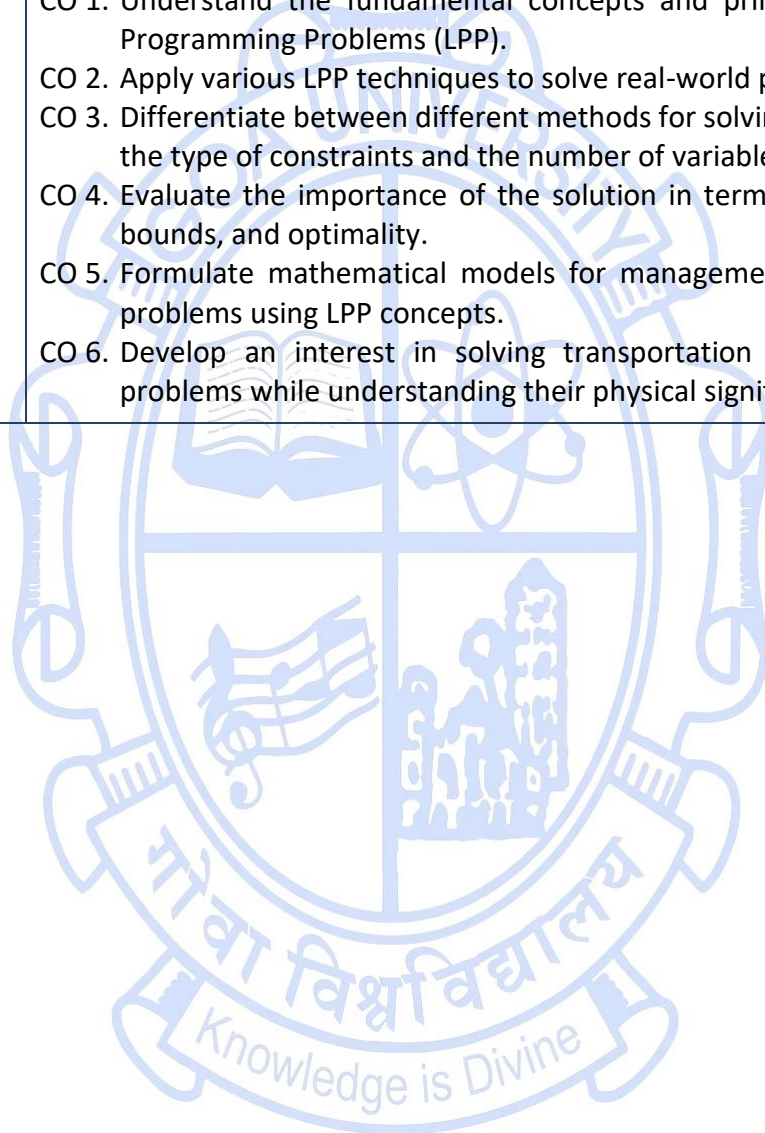
Pre-requisites for the Course:	NIL	
Course Objectives:	1. To introduce the fundamental concepts of and principles of data modeling. 2. To understand the role of data visualization in data analysis.	
Units	Content	No. of hours 75 (45T + 30P)
I	Data modeling - Meaning, significance, lifecycle, types of models. Relational model - Overview, keys in relational database, data anomalies, normalization and denormalization. ER modeling - ER diagrams, Cardinality constraints, crow's foot notation, recursive relationships, cross entity dependencies.	15T
II	Dimensional modeling - Star Schema vs. Snowflake Schema, fact tables, dimension tables, slow changing dimensions, data cubes and multidimensional modeling. NoSQL Data Modeling - NoSQL databases (Key-value stores, Document stores, Column-family stores, Graph databases), modeling for NoSQL vs. Relational databases, Schema design for MongoDB and Cassandra, scalability and consistency issues.	15T
III	Visualization - Meaning, components, significance of visualization in data analysis, design and cognitive principles, ethics and best practices. Data visualization types - right graph for right data. Visualization tools - Overview of popular data visualization tools (Tableau, Power BI etc.), basics of report and dashboard creation, web-based interactive data visualization platforms.	15T
IV	Practical Work	30P

	<p>Lab assignments/exercises based on a data modeling-visualisation platform like PowerBI/Tableau.</p> <p>Overview - Installation and set up, connecting to data.</p> <p>Modeling - Building a data model for a given business scenario.</p> <p>Charts - bar chart, line graph, pie chart, bubble chart, heat map.</p> <p>Calculations - sum, average, aggregation, custom calculations.</p> <p>Structuring data - sort, filter, pivot, custom filters, tooltips, color formatting.</p> <p>Dashboards - storytelling, design for different displays, interactive dashboards for a business scenario, dashboard sharing .</p>	
Pedagogy :	Lectures, Practical sessions, Assignments, Seminars, Presentations, Mini-projects or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading :	<ol style="list-style-type: none"> 1. Simsion, G., & Witt, G. (2004). Data modeling essentials. Elsevier. 2. Healy, K. (2018). Data visualization: a practical introduction. Princeton University Press. 3. Ware, C. (2019). Information visualization: perception for design. Morgan Kaufmann. 4. Ferrari, A., & Russo, M. (2016). Introducing Microsoft Power BI. Microsoft Press. 	
Course Outcomes :	<p>On completion of the course, students will be able to:</p> <p>CO 1. Understand data modeling principles and the usage of corresponding tools.</p> <p>CO 2. Build data models for real world business scenarios.</p> <p>CO 3. Understand the principles of effective data visualization.</p> <p>CO 4. Choose appropriate visualizations for different types of data.</p>	

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-203
Title of the Course : Linear Programming and Optimization
Number of Credits : 4
Effective from AY : 2024-25

Pre-requisites for the Course:	A basic understanding of linear algebra and calculus techniques for optimization, including matrix operations, optimality conditions. Familiarity with at least one software tool for numerical computations, such as Excel, R, MATLAB, Python, or an equivalent platform.	
Course Objectives:	1. To understand and apply problem-solving techniques both theoretically and graphically. 2. To develop the ability to analyze and interpret the results of mathematical models in real-world contexts. 3. To enhance problem-solving skills by applying optimization techniques to diverse scenarios.	
Units	Content	No. of hours 60
I	Linear Programming Introduction –Properties of Linear Programming-Basic assumptions-Mathematical formulation of Linear Programming. Limitations or constraints-Methods for the solution of LP	15T
II	Problem-Graphical analysis of LP-Graphical LP Maximization problem-Graphical LP Minimization problem. Simplex method and its variant: Simplex method, Artificial variable techniques- Two Phase Method; M-Charnes Method, Big M Method. Special cases in LPP.	15T
III	Duality: Definition of the dual problem, Primal-dual relationships, Economic Interpretation of Duality, Dual simplex Method.	15T
IV	Sensitivity analysis: Changes in cost and resource vector. Transportation and Assignment Models Introduction:Transportation problem - Balanced - Unbalanced - Methods of basic feasible solution Optimal solution-MODI method. Assignment problem-Hungarian Method.	15T
Pedagogy:	Lectures/ Practice Problems/ Assignments/ Presentations	

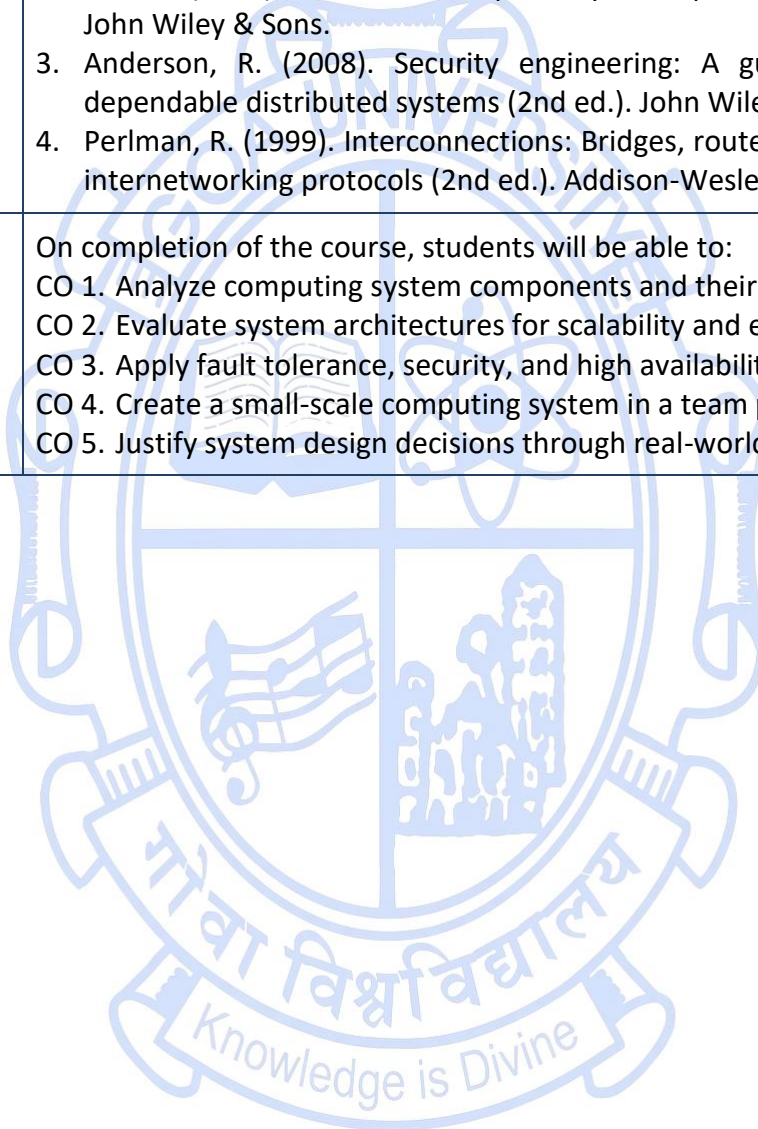
References/ Reading:	<ol style="list-style-type: none"> 1. Taha, H. A. (2008). Operations research: An introduction (8th ed.). Prentice Hall. 2. Hillier, F. S., & Lieberman, G. J. (2010). Introduction to operations research: Concepts and cases (9th ed.). Tata McGraw-Hill. 3. Hadley, G. (2002). Linear programming. Narosa (Reprint). 4. Ravindran, A., Phillips, D. T., & Solberg, J. J. (2005). Operations research: Principles and practice. John Wiley & Sons.
Course Outcomes:	<p>On completion of the course, students will be able to :</p> <p>CO 1. Understand the fundamental concepts and principles of Linear Programming Problems (LPP).</p> <p>CO 2. Apply various LPP techniques to solve real-world problems.</p> <p>CO 3. Differentiate between different methods for solving LPPs based on the type of constraints and the number of variables.</p> <p>CO 4. Evaluate the importance of the solution in terms of uniqueness, bounds, and optimality.</p> <p>CO 5. Formulate mathematical models for management and technical problems using LPP concepts.</p> <p>CO 6. Develop an interest in solving transportation and assignment problems while understanding their physical significance.</p>



Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-204
Title of the Course : Fundamentals of Computing Systems Design
Number of Credits : 4
Effective from AY : 2024-25

Pre-requisites for the Course:	Knowledge of Programming	
Course Objectives:	<ol style="list-style-type: none"> 1. To understand the key principles of computing system design, including application-level and system-level considerations. 2. To analyze and design modular, scalable, and efficient computing systems. 3. To explore performance estimation, fault tolerance, and security considerations in system design. 4. To gain hands-on experience through case studies and a team-based design project 	
Units	Content	No. of hours 60
I	Foundations of Computing System Design <ol style="list-style-type: none"> 1. Overview of computing systems; mobile computing; System design approach and key challenges 2. Applications; operating systems, and their interactions; Storage systems and file management; Computer architecture and hardware execution models 	15T
II	Modularity, Workloads & Performance Estimation <ol style="list-style-type: none"> 1. Modularity with clients and services; modularity with virtualization 2. Understanding different types of computing workloads 3. Performance estimation and optimization techniques 	15T
III	Networks, Fault Tolerance & Security <ol style="list-style-type: none"> 1. Networking fundamentals in system design; Networked applications and distributed systems 2. Designing fault-tolerant and resilient systems; Redundant system design principles 3. Enforcing information security in computing systems; Best practices for secure system architecture 	15T
IV	Case Studies & Project <ol style="list-style-type: none"> 1. Analysis of real-world systems such as YouTube, Twitter, Instagram, etc. 2. Design a small-scale computing system to address specific problems, so that the system is scalable, maintainable, and 	15T

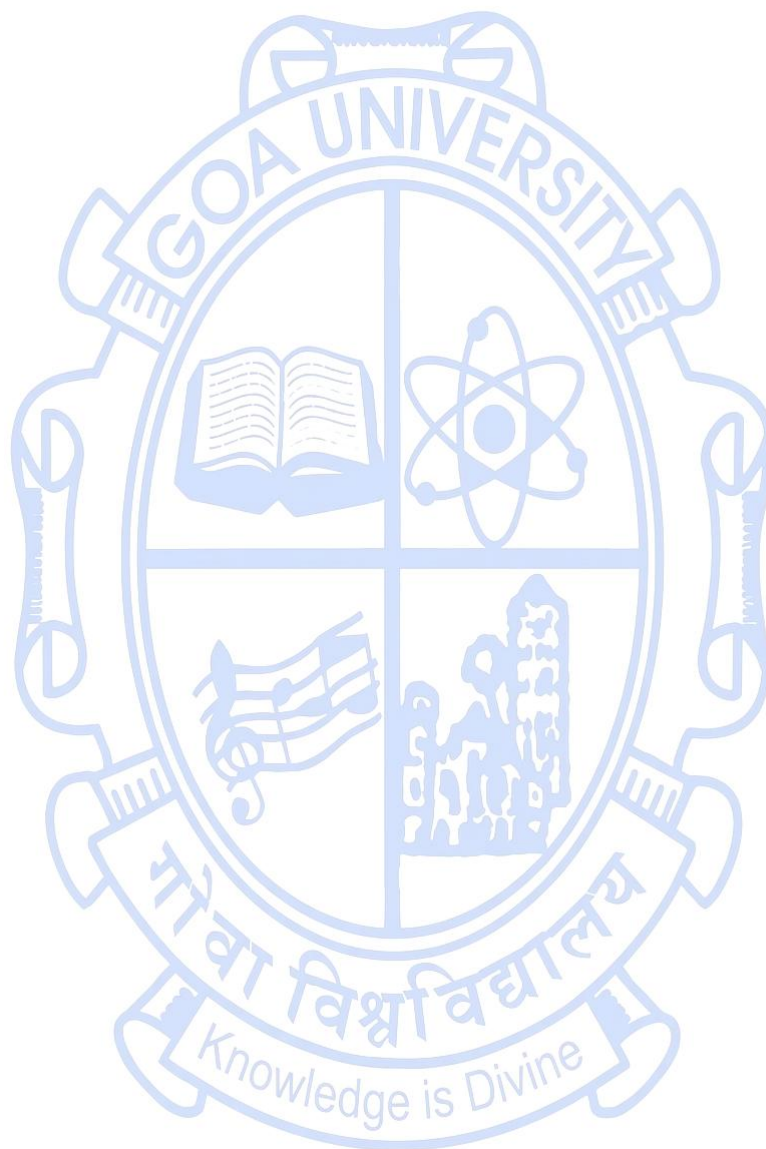
	fault tolerant	
Pedagogy:	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading :	<ol style="list-style-type: none"> 1. Saltzer, J., & Kaashoek, M. F. (2009). Principles of computer system design (1st ed.). Elsevier. 2. Jain, R. (1991). The art of computer systems performance analysis. John Wiley & Sons. 3. Anderson, R. (2008). Security engineering: A guide to building dependable distributed systems (2nd ed.). John Wiley & Sons. 4. Perlman, R. (1999). Interconnections: Bridges, routers, switches, and internetworking protocols (2nd ed.). Addison-Wesley 	
Course Outcomes :	<p>On completion of the course, students will be able to:</p> <p>CO 1. Analyze computing system components and their interactions.</p> <p>CO 2. Evaluate system architectures for scalability and efficiency.</p> <p>CO 3. Apply fault tolerance, security, and high availability principles.</p> <p>CO 4. Create a small-scale computing system in a team project.</p> <p>CO 5. Justify system design decisions through real-world case studies.</p>	



Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-205
Title of the Course : LINUX Fundamentals
Number of Credits : 2
Effective from AY : 2024-25

Pre-requisites for the Course:	NIL	
Course Objectives:	To introduce students to the Linux operating system environment and provide knowledge of basic Linux commands and shell scripting	
Units	Content	No. of hours 60
I	<ol style="list-style-type: none"> 1. Introduction - Linux Installation 2. Directory handling - cd, mkdir, rmdir, mv, pwd 3. File manipulation - cat, cp, ls, mv, rm, chmod, chown, find, cat, more, head, tail, cmp, wc, touch, pr 4. General purpose - date, history, man, who, whoami, uptime, finger, cal, uname, tree, bc, tar, zip 5. String manipulation - grep, egrep, cut, paste, tr, sort, rev, awk, sed 6. Process utilities - ps, pid, ppid, tty, time, kill, exit 7. Network utilities - ping, ifconfig, netstat, hostname, traceroute, telnet, ssh, mount, network configuration 	30P
II	<ol style="list-style-type: none"> 1. Shell Programming basics: Interactive shell script, positional parameters, operators, decision constructs, iterative constructs, Meta characters. 2. Shell programming for file manipulation - display contents, renaming based on extension, concatenation, find and display hidden files, counting the number of files, red-only write only files, file conversion to HTML, automate backup, deletion based on conditions, archive and compress log files. 3. Shell programming for user info - check if user exists, is logged in, password generation with criteria, generate reports of login activity, frequently used commands by the users. 	30P
Pedagogy:	Practical, tutorials, assignments, mini-projects, self-study	
References/ Reading:	<ol style="list-style-type: none"> 1. Das, S. (2006). UNIX, Concepts and Applications. Tata McGraw-Hill. 2. Glass, G., & Ables, K. (2003). UNIX for Programmers and Users. Pearson Education India. 3. Raymond, E. S. (2003). The art of Unix programming. Addison-Wesley Professional. 	

Course Outcomes:	On completion of the course, students will be able to : CO 1. Gain a deeper understanding of shell programming and its implications. CO 2. Implement efficient shell scripts on LINUX OS
-------------------------	--



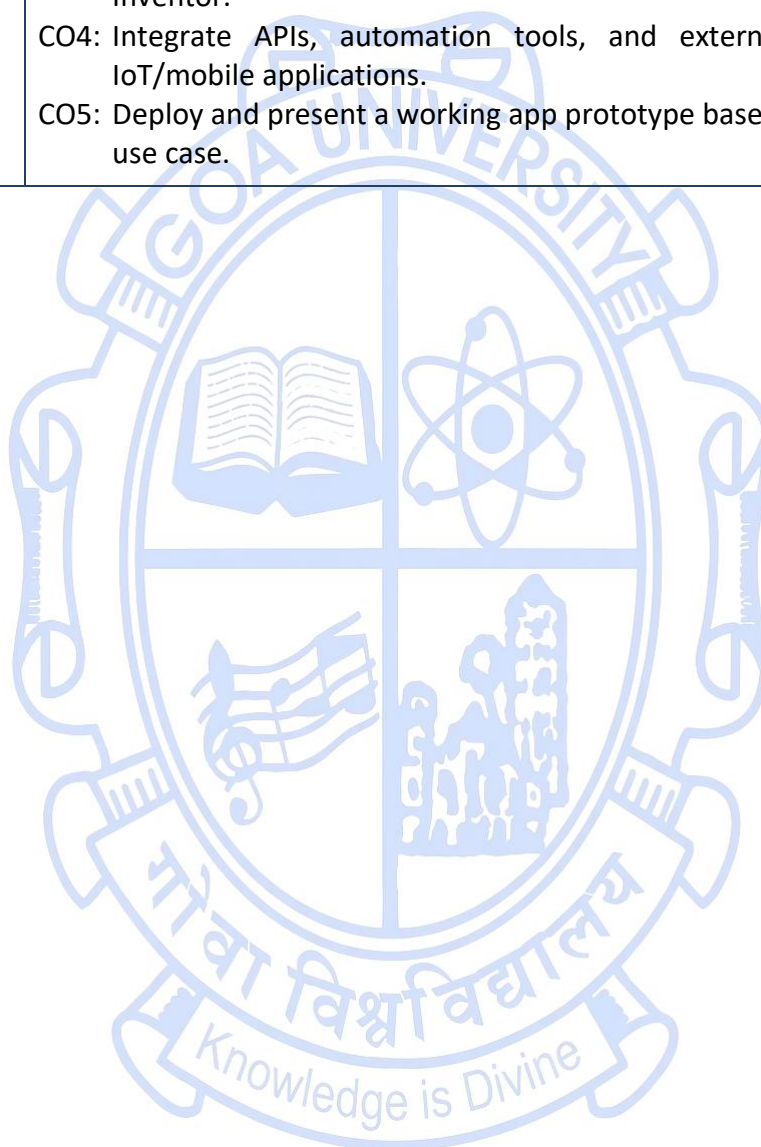
Minor VET Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-222
Title of the Course : Low Code App Development
Number of Credits : 4 (3T+1P)
Effective from AY : 2026-27

Pre-requisites for the Course:	NIL	
Course Objectives:	<ol style="list-style-type: none"> 1. To introduce learners to rapid application development using no-code and low-code platforms for web, mobile, and IoT applications. 2. To equip learners with the skills to design, build, and deploy functional applications without extensive coding knowledge. 3. To enable learners to integrate external services, automate processes, and develop real-world projects using RAD tools. 	
Units	Content	No. of hours (75) (45T+30P)
I	Introduction to Rapid Application Development (RAD) <ul style="list-style-type: none"> ● RAD principles and methodologies. ● Low-code vs. no-code development. ● Overview of web, mobile, and IoT applications. ● UI/UX principles for rapid development. ● Prototyping and wireframing tools. No-Code/Low-Code Web App Development <ul style="list-style-type: none"> ● Static vs. dynamic websites. ● Database-driven web applications. ● Using platforms for web development. ● Integrating APIs and automation tools. 	15T
II	No-Code/Low-Code Mobile App Development <ul style="list-style-type: none"> ● Mobile app & development lifecycle. ● Drag-and-drop platforms for mobile app development. ● Integrating mobile sensors and cloud services. ● Exporting and publishing mobile apps. 	15T
III	No-Code/Low-Code IoT App Development <ul style="list-style-type: none"> ● IoT application & development lifecycle. ● Platforms for IoT development. ● Integrating sensors and actuators. ● IoT app deployment considerations. 	15T

<p>IV</p>	<p>List of Suggested Lab Assignments</p> <ol style="list-style-type: none"> 1. No-Code Web Development with a platform (like WordPress); setting up a website; themes, plugins, and customization; forms, databases and automation tools 2. No-Code Web App Development with a platform (like Bubble, OutSystems, and Adalo); build simple web applications with drag-and-drop interfaces; connect web apps to external APIs 3. No-Code Mobile App Development with a platform (like MIT App Inventor); build a basic mobile app (UI, logic, and testing); integrate mobile sensors and cloud services 4. Export/publish mobile apps 5. Connecting mobile and web applications 6. IoT Application Development with a platform (like Pictoblox) for IoT and smart applications; connect sensors and microcontrollers; simple automation 7. IoT app deployment considerations. 8. UI/UX Design using design tools like Pencil/Figma/Sketch <p>Practical Work (Project)</p> <ul style="list-style-type: none"> ● Projection selection and planning, by brainstorming ideas, defining scope and feature and choosing the right platforms/tools for implementation ● Develop apps, using no-/low-code platforms, implement UI/logic/external integrations, test/debug applications ● Deployment/publish the app, for user feedback and peer review ● The project is ideally done in a group 	<p>30P</p>
<p>Pedagogy :</p>	<p>Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.</p>	
<p>References/ Reading :</p>	<ol style="list-style-type: none"> 1. Simon, P. (2022). Low-Code/No-Code: Citizen Developers and the Surprising Future of Business Applications. Wiley. 2. Helzle, S. (2022). Low-Code Application Development with Appian. Packt Publishing. 3. Murru, E. (2020). Hands-On Low-Code Application Development with Salesforce. Packt Publishing. 4. Sabin-Wilson, L. (2024). WordPress All-in-one for Dummies. John Wiley & Sons. 5. Reynolds, M. (2021). No-Code AI and Machine Learning: Building AI Applications Without Coding. Apress. 6. STEMpedia. (2023). Pictoblox for Beginners: AI, IoT, and Robotics with 	

	No-Code Programming. STEMpedia Press.
Course Outcomes:	<p>On completion of the course, students will be able to :</p> <p>CO1: Describe the principles and advantages of rapid application development using no-code and low-code tools.</p> <p>CO2: Demonstrate the ability to create web applications using no-code/low-code platforms.</p> <p>CO3: Develop functional mobile applications using platforms like MIT App Inventor.</p> <p>CO4: Integrate APIs, automation tools, and external services into IoT/mobile applications.</p> <p>CO5: Deploy and present a working app prototype based on a real-world use case.</p>



Semester V

Major Courses

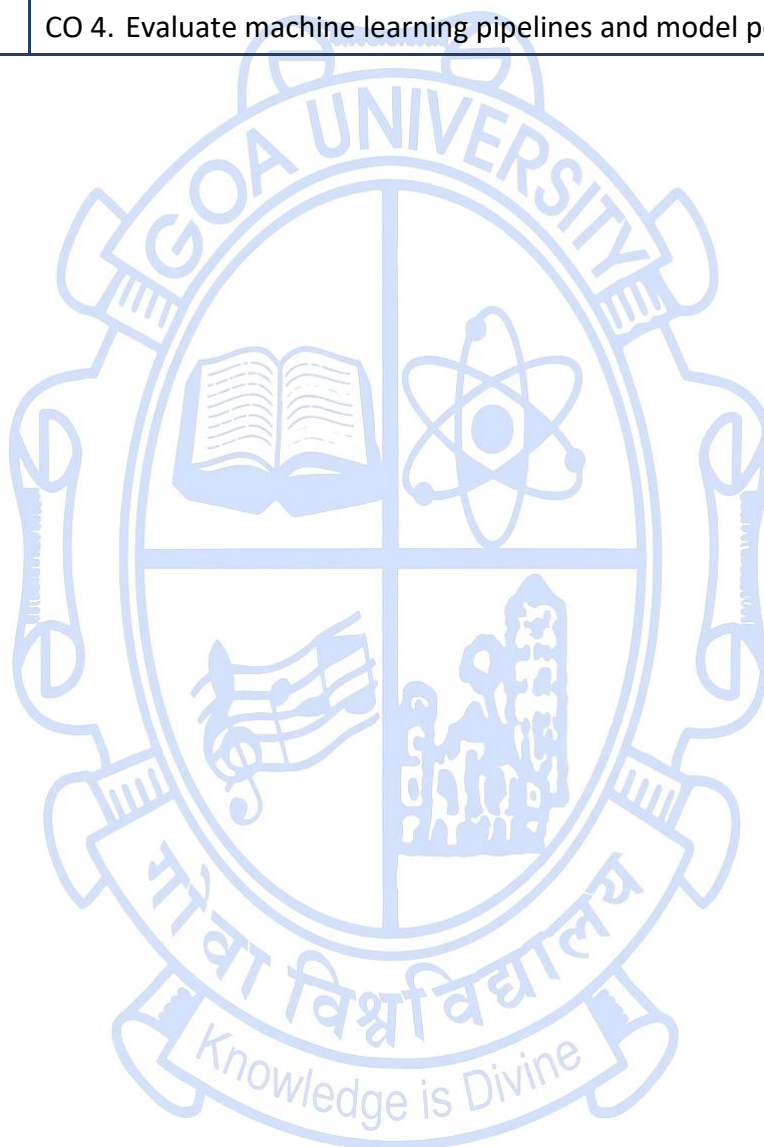
Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-300
Title of the Course : Introduction to Machine Learning
Number of Credits : 4 (2T+2P)
Effective from AY : 2024-25

Pre-requisites for the Course :	IDS-100 Mathematical Foundations,	
Course Objectives :	The objectives of this course are to: <ol style="list-style-type: none"> 1. To introduce fundamental machine learning concepts, including concept learning, decision trees, and random forests. 2. To apply supervised learning techniques such as regression, classification, Naïve Bayes, SVM, and neural networks to real-world problems. 3. To analyze clustering techniques and PCA for pattern discovery and dimensionality reduction. 4. To evaluate machine learning models and pipelines using appropriate performance metrics. 	
Units	Content	No. of hours 90 (30T+60P)
I	<p>Unit I: Concept Learning and Supervised Learning</p> <p>Introduction to Concept Learning: Concept learning, hypothesis space, version space, decision tree learning, random forest algorithm.</p> <p>Linear Models for Regression: Introduction to regression, linear basis function models, least squares method, geometry of least squares, regularized least squares, bias-variance decomposition, introduction to Bayesian linear regression.</p> <p>Linear Models for Classification: Introduction to classification problems, discriminant functions, two-class and multiclass classification problems, Fisher's linear discriminant, Naïve Bayes classifier (concept and working principle), binary logistic regression, multiclass logistic regression, comparison of regression and classification models.</p> <p>Introduction to Neural Networks: Biological inspiration of neural networks, perceptron algorithm, basic feed-forward</p>	15T

	neural network structure, idea of network training, gradient descent optimization, error backpropagation.	
II	<p>Unit II: Kernel Methods, Unsupervised Learning, and Ensemble Methods</p> <p>Support Vector Machines: Introduction to sparse kernel machines, maximum margin classifiers, intuition behind separating hyperplanes, linear SVM, SVM for regression, comparison of SVM and logistic regression.</p> <p>Unsupervised Learning and Clustering: Introduction to unsupervised learning, k-means clustering algorithm, applications of clustering, mixture of Gaussians (conceptual), expectation–maximization (EM) algorithm (high-level understanding).</p> <p>Dimensionality Reduction: Need for dimensionality reduction, principal component analysis (PCA), maximum variance intuition, applications of PCA, PCA for high-dimensional data.</p>	15T
III	<p>List of Suggested Lab Assignments</p> <p>Experiment 1: EDA and Data Visualization Analyze and visualize datasets (e.g., Iris, Titanic) using tools such as Pandas and Matplotlib to understand data distribution, relationships between variables, and basic preprocessing.</p> <p>Experiment 2: k-Nearest Neighbors (k-NN) Implement k-nearest neighbors classification on datasets such as Iris, tune the value of k, and evaluate performance using accuracy and confusion matrix.</p> <p>Experiment 3: Linear Regression Predict continuous target variables using linear regression on datasets such as Boston Housing or simple regression datasets, and evaluate performance using mean squared error and R^2 score.</p> <p>Experiment 4: Logistic Regression Perform binary classification using logistic regression on datasets such as Breast Cancer, and evaluate the model using accuracy, precision, recall, F1-score, and ROC curve.</p> <p>Experiment 5: Decision Trees and Random Forests Build decision tree and random forest models on datasets such as Titanic, compare their performance, study overfitting, and analyze feature importance.</p>	30P

IV	<p>List of Suggested Lab Assignments</p> <p>Experiment 6: Naïve Bayes Classification Implement Naïve Bayes classifier on datasets such as Iris or text-based spam datasets, and evaluate classification performance using accuracy and confusion matrix.</p> <p>Experiment 7: Support Vector Machines (SVM) Classify data using support vector machines on datasets such as Iris, experiment with linear kernels, and compare SVM performance with logistic regression.</p> <p>Experiment 8: Clustering using k-Means Apply k-means clustering on datasets such as Mall Customer Segmentation, visualize clusters, and evaluate clustering quality using inertia or silhouette score.</p> <p>Experiment 9: Principal Component Analysis (PCA) Apply principal component analysis on high-dimensional datasets such as Iris or MNIST (subset), reduce dimensionality, and visualize the transformed data.</p> <p>Experiment 10: Neural Networks Build and train a basic feed-forward neural network for classification using datasets such as MNIST (subset) or Iris, and evaluate model performance.</p>	30P
Pedagogy :	Lectures/Hands-on assignments/Flipped classroom/Mini project/Lab assignment	
References/ Reading :	<ol style="list-style-type: none"> 1. Alpaydin, E. (2020). Introduction to machine learning. MIT press, 4th Edition. 2. Mitchell, T. M., (1997). Machine learning. New York: McGraw-hill. 3. Bishop, C. M. (2006). Pattern recognition and machine learning: springer New York.. 4. Flach, P. (2012). Machine learning: the art and science of algorithms that make sense of data. Cambridge university press. 5. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. 6. Hart, Peter E., David G. Stork, and Richard O. Duda.(2000) Pattern classification. Hoboken: Wiley, 2000. 7. Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media 8. Starmer, J. (2023). The StatQuest illustrated guide to machine 	

	learning. StatQuest.
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. Understand basic machine learning concepts, including concept learning, decision trees, and random forests.</p> <p>CO 2. Apply regression, classification, Naïve Bayes, SVM, and neural networks to solve real-world problems.</p> <p>CO 3. Analyze clustering and PCA results to identify patterns and reduce dimensionality.</p> <p>CO 4. Evaluate machine learning pipelines and model performance.</p>



Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-301
Title of the Course : Data Mining and Warehousing
Number of Credits : 4(2T+2P)
Effective from AY : 2024-25

Pre-requisites for the Course:	IDS-102 Data Management	
Course Objectives:	The objectives of this course are to: <ol style="list-style-type: none"> 1. Introduce fundamental concepts, architectures, and models of data warehousing systems. 2. Develop an understanding of data preprocessing, integration, and OLAP operations for analytical processing. 3. Enable students to apply core data mining techniques such as association analysis, classification, clustering, and prediction. 4. Equip students with the ability to select appropriate data mining methods for real-world data analysis problems. 	
Units	Content	No. of hours 90 (30T+60P)
I	<p>Introduction: Data sources, Applications of Data Mining, KDD Process, Issues and Challenges. The need for Data Warehousing, OLTP vs OLAP.</p> <p>Data preprocessing: Data preprocessing overview, Data cleaning, Data integration, Multidimensional data model, Schemas for multidimensional databases - Star schema, Snowflake schema, OLAP concepts and operations</p> <p>Association Analysis: Different methods for mining association rules in transaction based databases. Illustration of confidence and support. Multidimensional and multilevel association rules. Classification of association rules. Discussion on few association rule algorithms e.g. Apriori, frequent pattern growth etc.</p>	15T
II	<p>Classification and Predictions: Different Classification algorithms, including C4.5, CART etc., use of genie index, decision tree induction, Bayesian classification, neural network technique of back propagation, fuzzy set theory and genetic algorithms.</p> <p>Clustering: Partition based clustering, Hierarchical clustering, model based clustering for continuous and discrete data. Discussion on scalability of clustering algorithms. Parallel approaches for clustering.</p>	15T
III	List of Suggested Lab Assignments	30P

	<ol style="list-style-type: none"> 1. Design a star schema for a given business dataset identifying fact and dimension tables. 2. Create and populate data warehouse tables using SQL. 3. Implement ETL/ELT operations (extraction, transformation, and loading) using Python/SQL. 4. Perform OLAP operations such as roll-up, drill-down, slice, and dice using SQL queries. 5. Preprocess data by handling missing values, normalization, and transformation using Python. 	
IV	<p>List of Suggested Lab Assignments</p> <ol style="list-style-type: none"> 1. Implement the Apriori algorithm to generate frequent itemsets and association rules. 2. Implement decision tree classification and evaluate model accuracy. 3. Implement Naïve Bayes classification and compare results with the decision tree. 4. Perform k-means clustering and analyze cluster formation. 5. Apply model evaluation techniques such as confusion matrix and cross-validation on mining models. 	30P
Pedagogy :	Lectures/ Tutorials/Hands-on assignments/Self-study/Lab Assignments	
References/ Reading :	<ol style="list-style-type: none"> 1. Han, J., Pei, J., & Tong, H. (2022). <i>Data mining: concepts and techniques(4th ed.)</i>. Morgan Kaufmann. 2. Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to data mining. Pearson Education. 3. Margaret, H. (2006). Data Mining-“Introductory and Advanced Concepts”. 4. Pujari, A. K. (2001). <i>Data mining techniques</i>. Universities press. 	
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. Explain the concepts, architecture, and modeling techniques of data warehousing systems.</p> <p>CO 2. Apply data preprocessing and OLAP operations for analytical and decision-support tasks.</p> <p>CO 3. Implement core data mining techniques such as association analysis, classification, and clustering.</p> <p>CO 4. Evaluate data mining models to extract meaningful insights from real-world datasets.</p>	

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-302
Title of the Course : Business Data Management
Number of Credits : 4T
Effective from AY : 2024-25

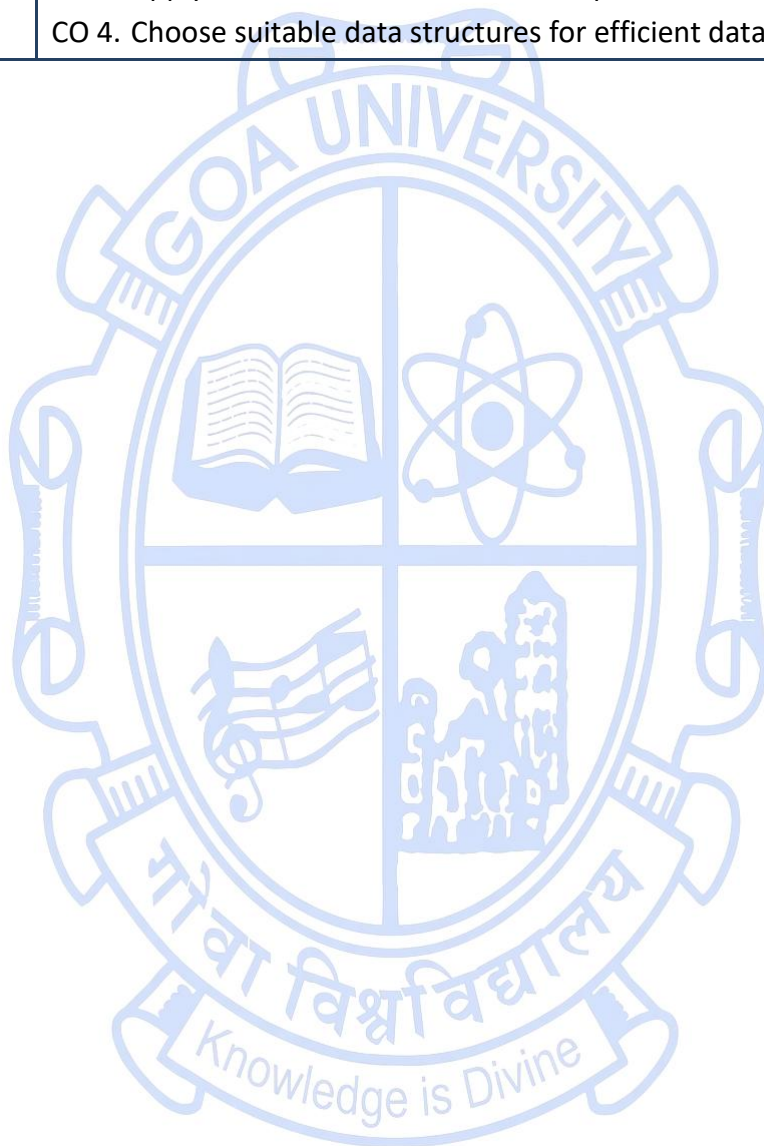
Pre-requisites for the Course :	NIL	
Course Objectives :	The objectives of this course are to: <ol style="list-style-type: none"> 1. Provide a comprehensive understanding of business data, its organization, and its role in modern enterprises. 2. Develop knowledge of database concepts and data management techniques used in business environments. 3. Equip learners with the ability to manage, query, and maintain business data using standard tools and practices. 4. Enable informed decision-making by applying data management principles to real-world business scenarios. 	
Units	Content	No. of hours 60T
I	Fundamentals of Business Data: Introduction to data and information. Role of data in business operations and strategy. Types of business data: structured, semi-structured, and unstructured. Sources of business data: internal, external, and third-party sources Data lifecycle: creation, storage, usage, archival, and disposal Basics of data-driven decision-making Overview of business data ecosystems	15T
II	Data Modeling and Organization : Data organization concepts and file-based systems. Introduction to data models: hierarchical, network, relational, and object-oriented models. Entity-Relationship (ER) modeling: entities, attributes, relationships, and constraints. Normalization concepts and functional dependencies. Data integrity, consistency, and redundancy issues. Metadata management and data dictionaries. Business use cases of data modeling	15T
III	Database Management Systems for Business: Introduction to Database Management Systems (DBMS). Components and architecture of DBMS. Relational database	15T

	<p>concepts: tables, keys, and relationships.</p> <p>Structured Query Language (SQL): Data Definition Language (DDL). Data Manipulation Language (DML). Basic queries, joins, and subqueries. Transaction management and concurrency control</p> <p>Applications of DBMS in business functions (HR, finance, marketing, supply chain)</p>	
IV	<p>Business Data Management Practices and Governance:</p> <p>Data quality management: accuracy, completeness, consistency, and timeliness. Data security concepts: access control, authentication, and authorization. Privacy, ethical, and legal issues in business data management. Data backup, recovery, and disaster management. Introduction to data governance frameworks. Compliance standards and regulations affecting business data</p> <p>Emerging trends: cloud databases, big data, and data-driven enterprises</p>	15T
Pedagogy :	Lectures/ Tutorials/Hands-on assignments/Self-study	
References/ Reading :	<ol style="list-style-type: none"> 1. Silberschatz, A., Korth, H. F., & Sudarshan, S. (2019). Database system concepts (7th ed.). McGraw-Hill Education. 2. Coronel, C., Morris, S., & Rob, P. (2019). Database systems: Design, implementation, and management (13th ed.). Cengage Learning. 3. Connolly, T., & Begg, C(2015). Database Systems: A Practical Approach to Design, Implementation, and Management(6th ed.). Pearson Education. 	
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. Understand and explain the concepts, types, and importance of data in business organizations.</p> <p>CO 2. Design and organize business data using appropriate data models and structures.</p> <p>CO 3. Apply database management concepts and query techniques for effective business data handling.</p> <p>CO 4. Evaluate data quality, security, and governance practices in business contexts.</p>	

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-303
Title of the Course : Data Structures
Number of Credits : 2T
Effective from AY : 2024-25

Pre-requisites for the Course:	IDS-111 Problem Solving and Program Design IDS-142 Programming in Python	
Course Objectives :	The objectives of this course are to: 1. Introduce fundamental linear and non-linear data structures. 2. Develop the ability to analyze time and space complexity. 3. Enable students to implement data structures using programming languages. 4. Equip students with skills to select appropriate data structures for data science problems.	
Units	Content	No. of hours 30T
I	Unit I - Linear Data Structures Introduction to Data Structures: Abstract Data Types (ADT), Time and Space Complexity (Big-O notation) Arrays and Strings: Representation and operations, Applications in data processing Linked Lists: Singly and Doubly Linked Lists. Insertion, deletion, traversal Stacks and Queues: Stack operations and applications. Queue types: Simple Queue, Circular Queue	15T
II	Unit II – Non-Linear Data Structures Trees: Tree terminology and properties. Binary Trees and Binary Search Trees. Tree traversals (Inorder, Preorder, Postorder) Heaps: Min-Heap and Max-Heap, Priority Queues and Applications Graphs: Graph representation (Adjacency Matrix & List). Graph traversal techniques: BFS and DFS. Applications in network and data analysis	15T
Pedagogy :	Lectures/ Tutorials/Hands-on assignments/demonstrations/Self-study	
References/ Reading :	1. Goodrich, M. T., Tamassia, R., & Goldwasser, M. H. (2013). Data structures and algorithms in Python. Wiley. 2. Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). Introduction to algorithms (4th ed.). MIT Press. 3. Miller, B. N., & Ranum, D. L. (2014). Problem solving with algorithms	

	<p>and data structures using Python. Franklin, Beedle & Associates.</p> <p>4. Thareja, R. (2018). Data structures using C (2nd ed.). Oxford University Press.</p>
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. Explain fundamental concepts of data structures and ADTs.</p> <p>CO 2. Implement and apply linear data structures using programming languages.</p> <p>CO 3. Apply non-linear data structures to problem-solving.</p> <p>CO 4. Choose suitable data structures for efficient data science solutions.</p>

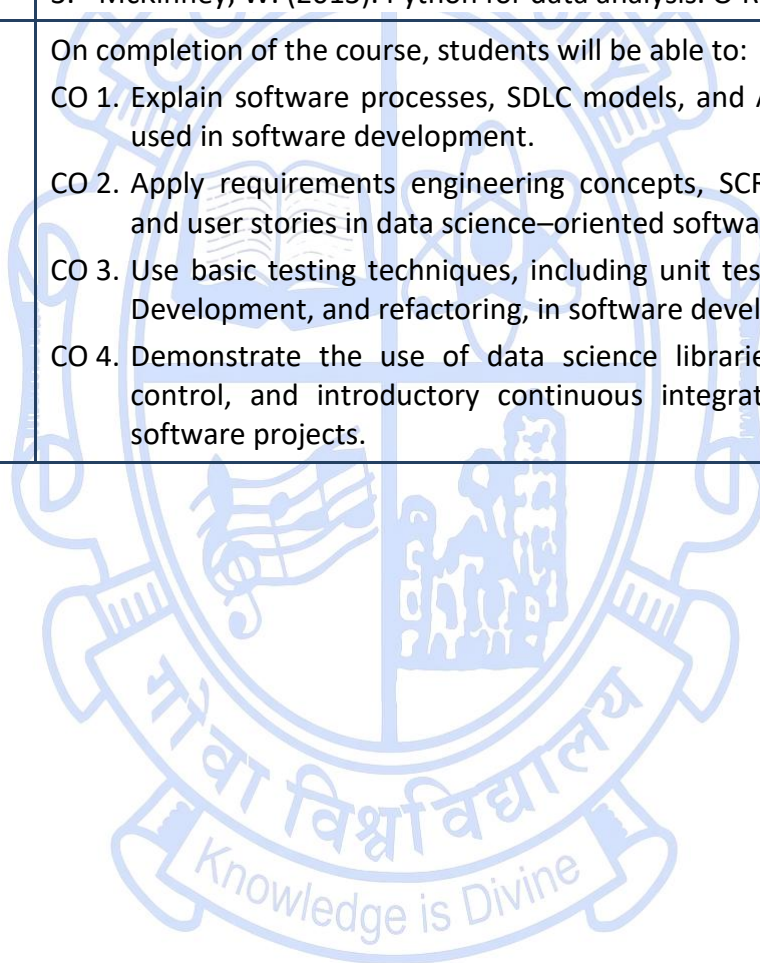


Minor VET Course

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-321
Title of the Course : Software Engineering
Number of Credits : 4 (3T+1P)
Effective from AY : 2024-25

Pre-requisites for the Course :	IDS-142 Programming in Python	
Course Objectives :	The objectives of this course are to: <ol style="list-style-type: none"> 1. To introduce fundamental software engineering concepts, software processes, SDLC models, and Agile approaches. 2. To develop an understanding of requirements engineering, SCRUM framework, and user stories for data science applications. 3. To familiarize students with basic testing practices, including unit testing, Test Driven Development, and refactoring. 4. To provide exposure to software development practices for data science, including use of frameworks, APIs, data handling, version control, and continuous integration. 	
Units	Content	No. of hours 75 (45T+30P)
I	Software Engineering: Introduction to software processes and SDLC. Agile approaches to software engineering. Requirements engineering for data science applications. Overview of SCRUM framework and user stories.	15T
II	Introduction to Test Driven Development: Unit testing. Data driven development and refactoring.	15T
III	Software Development Practices for Data Science: Use of software frameworks and APIs for data science applications.(such as NumPy, Pandas). Basics of data handling and data pipelines. Configuration management and version control. Introduction to continuous integration and automated software engineering.	15T
IV	List of Suggested Lab Assignments <ol style="list-style-type: none"> 1. Git Basics for Data Science Projects 2. Introduction to Unit Testing in Python 3. Test Driven Development and Refactoring 4. Data Handling using Python Libraries 	30P

	<ul style="list-style-type: none"> 5. Building Data Science Modules using Frameworks 6. API Usage in Data Science Applications 7. Functional Testing of Data Science Applications 8. Performance and API Testing 9. Introduction to Continuous Integration 	
Pedagogy :	Lectures/ Tutorials/ Hands-on assignments/Self-study/ Flipped classroom/ Lab assignments	
References/ Reading :	<ul style="list-style-type: none"> 1. Sommerville, I. (2011). Software engineering (9th ed.). Pearson Education India. 2. Pressman, R. S. (2005). Software engineering: A practitioner's approach (6th ed.). Palgrave Macmillan. 3. McKinney, W. (2013). Python for data analysis. O'Reilly Media. 	
Course Outcomes :	<p>On completion of the course, students will be able to:</p> <p>CO 1. Explain software processes, SDLC models, and Agile approaches used in software development.</p> <p>CO 2. Apply requirements engineering concepts, SCRUM framework, and user stories in data science-oriented software projects.</p> <p>CO 3. Use basic testing techniques, including unit testing, Test Driven Development, and refactoring, in software development.</p> <p>CO 4. Demonstrate the use of data science libraries, APIs, version control, and introductory continuous integration practices in software projects.</p>	



Semester VI

Major Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-304
Title of the Course : Cloud Computing
Number of Credits : 4 (2T+2P)
Effective from AY : 2024-25

Pre-requisites for the Course :	IDS-205 LINUX Fundamentals	
Course Objectives :	<p>The objectives of this course are to:</p> <ol style="list-style-type: none"> 1. Develop understanding of cloud service models, and infrastructure provisioning. 2. Enable students to utilize virtualization and containerization technologies like Docker for efficient data science workflows. 3. Develop skills in designing scalable data pipelines using cloud storage, managed databases, and serverless architectures. 4. Familiarize learners with cloud security frameworks and strategies for deploying production-ready Data Science applications. 	
Units	Content	No. of hours 90 (30T+60P)
I	<p>Computer Network Fundamentals TCP/IP suite, IP addresses, DNS, Network Metrics – Latency, Resource Usage.</p> <p>Cloud Fundamentals & Containerization Overview: Cloud Evolution, NIST Model, Service Models (IaaS, PaaS, SaaS). Virtualization vs. Containerization: Virtual Machines (VMs) vs. Containers. Use of containerization for Data Science. Compute Fundamentals – Provisioning Virtual Machines for Data Processing. Container Fundamentals: Docker architecture, Dockerfiles, and Container Registries.</p> <p>Cloud Storage & Data Management Storage Paradigms: Block Storage vs. Object Storage provisioning for Data Lakes. Cloud Databases: Managed Relational Databases vs. Managed NoSQL provisioning.</p>	15T

	Data Lifecycle: Tiered storage (Hot/Cool/Archive) for cost-effective dataset management.	
II	<p>Serverless data processing & Big Data</p> <p>Serverless Computing: Function-as-a-Service. Event-driven architecture for automated ETL pipelines.</p> <p>Cloud Data Warehousing and Analytics: Massively Parallel Processing (MPP) architecture, Decoupled storage and compute, Serverless query execution, Columnar storage and partitioning, ELT data pipelines, Auto-scaling and cost governance, Data Lakehouse.</p> <p>Data Analytics as a Service: Overview of cloud-native tools for large-scale data querying without managing infrastructure.</p> <p>Security and MLOps basics</p> <p>Cloud Security: Shared Responsibility Model. Identity and Access Management (IAM) for securing datasets.</p> <p>ML Deployment Patterns: PaaS offerings for hosting ML dashboards.</p>	15T
III	<p>List of Suggested Lab Assignments</p> <p>Suggested Lab assignments demonstrating application of following concepts:</p> <ol style="list-style-type: none"> 1. Virtual Machine provisioning and configuration of security groups/firewalls for remote access (SSH/RDP) and data science environment setup. 2. Containerization of data science applications using Docker: creating Dockerfiles, building images, and managing local containers and pushing them to Cloud Registries. 3. Cloud Object Storage implementation: creating buckets/containers and applying Lifecycle Management policies (Hot/Cool/Archive) for cost optimization. 4. Managed Cloud Database operations: provisioning and connecting to managed Relational (SQL) or NoSQL databases. 	30P
IV	<p>List of Suggested Lab Assignments</p> <p>Suggested Lab assignments demonstrating application of following concepts:</p> <ol style="list-style-type: none"> 1. Serverless computing: developing functions (FaaS) to create event-driven ETL pipelines triggered by storage events. 2. Serverless Data Analytics: performing SQL queries on large-scale datasets directly from object storage 	30P

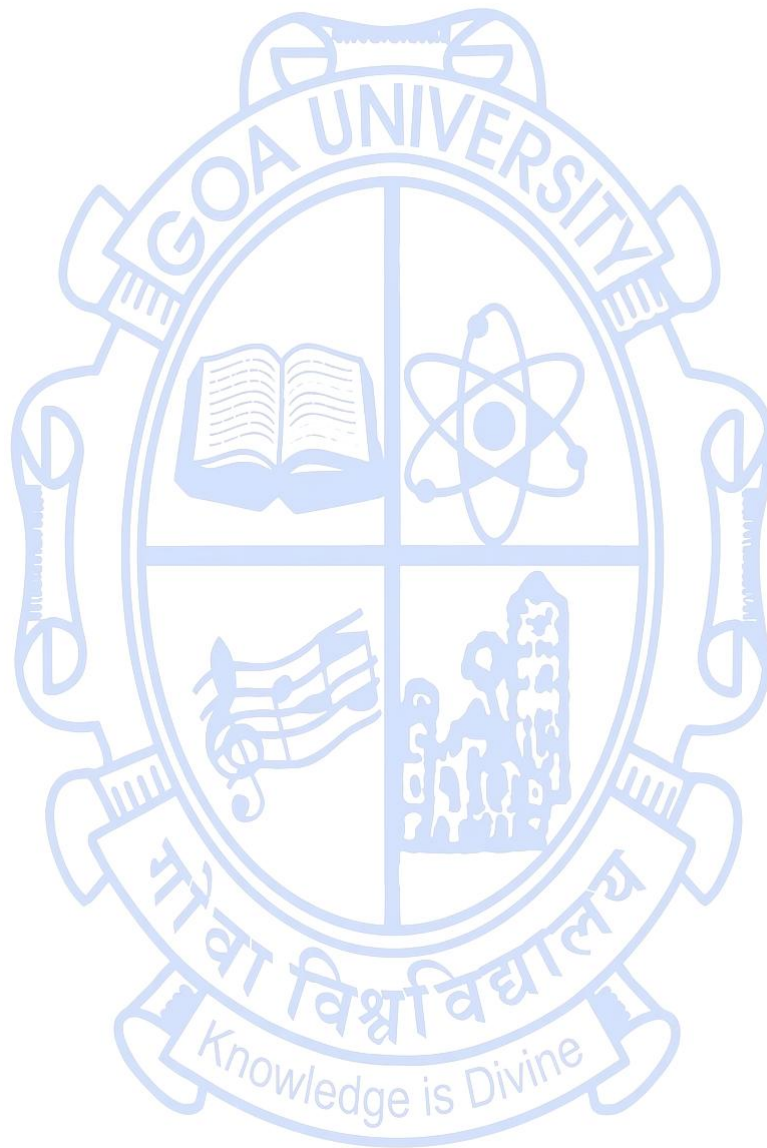
	<p>without infrastructure provisioning.</p> <ol style="list-style-type: none"> 3. Cloud Security implementation: configuring Identity and Access Management (IAM) users, roles, and policies to enforce least-privilege access. 4. ML Deployment patterns: hosting machine learning dashboards or APIs using Platform-as-a-Service (PaaS) web offerings. 	
Pedagogy :	Lectures/tutorials/ assignments/ mini-projects/ PPT presentations/ case studies/ class discussions/practicals.	
References/ Reading :	<ol style="list-style-type: none"> 1. Hwang, K., Fox, G. C., & Dongarra, J. J. (2011). Distributed and cloud computing: From parallel processing to the Internet of Things (1st ed.). Morgan Kaufmann. 2. Buyya, R., Broberg, J., & Goscinski, A. M. (2013). Cloud computing: Principles and paradigms. Wiley. 3. Davis, C. (2019). Cloud native patterns: Designing change-tolerant systems. O'Reilly Media. 4. Laszewski, T., Arora, K., Farr, E., & Zonooz, P. (2018). Cloud native architectures: Design high-availability and cost-effective applications for the cloud. Packt Publishing. 5. Morris, K. (2016). Infrastructure as code: Dynamic systems for the cloud age. O'Reilly Media. <p>Web resources:</p> <ol style="list-style-type: none"> 1. AWS. "AWS documentation." AWS Documentation, AWS, https://docs.aws.amazon.com/. Accessed 6 January 2026. 2. Google Cloud Platform. "Google Cloud Documentation." Google Cloud, https://cloud.google.com/docs. Accessed 6 January 2026. 3. Microsoft. "Azure documentation." Microsoft Learn, https://learn.microsoft.com/en-us/azure. Accessed 6 January 2026. 	
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. Analyze the service paradigms, deployment models, and architectural principles of public and private cloud computing.</p> <p>CO 2. Design scalable data pipelines using cloud-native storage, managed databases, and serverless computing.</p> <p>CO 3. Implement robust security controls and identity management policies to protect sensitive cloud resources.</p> <p>CO 4. Deploy machine learning models and automated infrastructure.</p>	

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-305
Title of the Course : Business Analytics and Intelligence
Number of Credits : 4 (2T+2P)
Effective from AY : 2024-25

Pre-requisites for the Course :	NIL	
Course Objectives :	The objectives of this course are to: <ol style="list-style-type: none"> 1. Introduce the concepts, techniques, and tools of business analytics and business intelligence used in organizations. 2. Develop analytical thinking for transforming business data into meaningful insights. 3. Enable learners to apply descriptive, predictive, and prescriptive analytics for business decision-making. 4. Familiarize students with business intelligence systems, dashboards, and reporting mechanisms. 	
Units	Content	No. of hours 90 (30T+60P)
I	Introduction to Business Analytics and Intelligence Overview of business analytics and business intelligence Evolution of analytics in business decision-making Types of analytics: descriptive, diagnostic, predictive, and prescriptive Data-driven organizations and analytics maturity models Business problems and analytical problem formulation Role of analytics in functional areas: marketing, finance, HR, and operations Introduction to analytics tools and platforms	15T
II	Data Management and Analytical Foundations Business data sources and data collection methods Data preprocessing concepts: data cleaning, integration, and transformation Descriptive analytics techniques and applications Statistical foundations for analytics Data visualization principles and best practices Key Performance Indicators (KPIs) and performance measurement systems Case discussions on analytics-driven decision-making	15T

III	<p>List of Suggested Lab Assignments: Predictive and Prescriptive Analytics</p> <p>Hands-on data preprocessing using analytics tools</p> <p>Exploratory data analysis using real-world business datasets</p> <p>Regression analysis and forecasting techniques</p> <p>Classification methods and model evaluation</p> <p>Clustering techniques for business segmentation</p> <p>Introduction to optimization models</p> <p>Simulation techniques for business decision-making</p> <p>Case-based analytics problem solving</p> <p>Interpretation, validation, and presentation of analytical results</p> <p>Practice sessions using industry-relevant datasets</p>	30P
IV	<p>List of Suggested Lab Assignments: Business Intelligence Tools and Applications</p> <p>Hands-on use of Business Intelligence (BI) tools</p> <p>OLAP operations and multidimensional data analysis</p> <p>Dashboard design principles and interactive visualizations</p> <p>Development of managerial reports and scorecards</p> <p>Integration of analytics outputs into BI dashboards</p> <p>Data governance, ethics, and compliance in BI systems</p> <p>Mini-project: end-to-end business analytics and intelligence solution</p> <p>Project presentation and evaluation</p>	30P
Pedagogy :	Lectures/tutorials/ assignments/ mini-projects/ PPT presentations/ case studies/ class discussions./practicals	
References/ Reading :	<ol style="list-style-type: none"> 1. Sharda, R., Delen, D., & Turban, E. (2020). Business intelligence, analytics, and data science: A managerial perspective (4th ed.). Pearson Education. 2. Davenport, T. H., & Harris, J. G. (2017). Competing on analytics: The new science of winning (2nd ed.). Harvard Business School Press. 3. Marr, B. (2016). Data strategy: How to profit from a world of big data, analytics and the Internet of Things. Kogan Page. 	
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. Explain the role of business analytics and intelligence in organizational strategy and performance management.</p> <p>CO 2. Apply analytical techniques to analyze and interpret business data.</p> <p>CO 3. Use business intelligence concepts and tools for reporting,</p>	

	visualization, and decision support. CO 4. Evaluate analytics-driven solutions to solve real-world business problems.
--	--

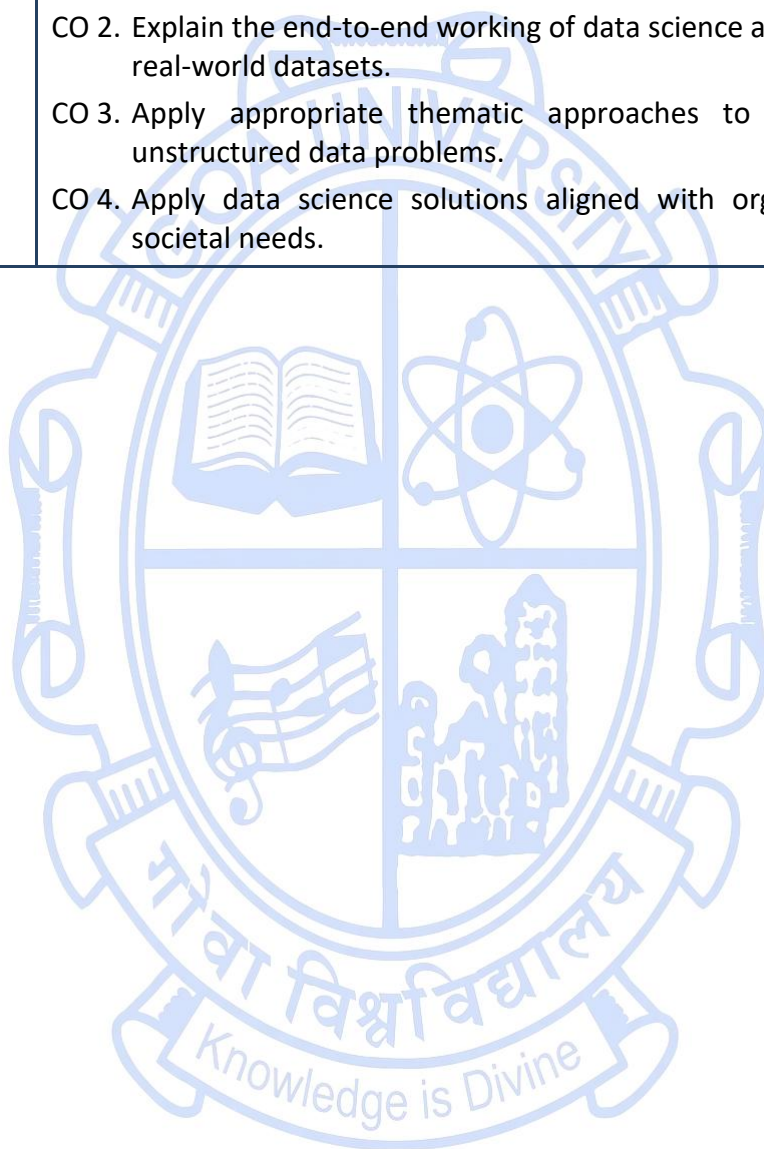


Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-306
Title of the Course : Thematic Ideas in Data Science
Number of Credits : 4T
Effective from AY : 2024-25

Pre-requisites for the Course :	IDS-200 Introduction to Data Science, IDS-300 Introduction to Machine Learning	
Course Objectives :	The objectives of this course are to: <ol style="list-style-type: none"> 1. To introduce and explain the fundamental thematic ideas in data science applications. 2. To analyze the working principles of machine learning, deep learning, and analytical models in real-world scenarios. 3. To apply thematic ideas in data science to solve domain-specific business problems. 4. To understand ethical, scalable, and data-driven solutions using modern data science tools. 	
Units	Content	No. of hours 60T
I	Evidence-based decision making and analytics lifecycle: Overview of Thematic Ideas in Data Science, Data-Driven vs Intuition-Based Decision Making; Types of Analytics: Descriptive, Diagnostic, Predictive, Prescriptive, Data Science Lifecycle; Cross-Industry Standard Process for Data Mining (CRISP-DM), Data Collection, Cleaning, and Exploratory Data Analysis; Visualization for Insight Generation; Ethics, Bias, Privacy, and Responsible AI.	15T
II	Predictive Analytics and Segmentation: Prediction, pattern discovery, and cost optimization; Employee Attrition Prediction (Human Resources); <ul style="list-style-type: none"> ● Available employee dataset is used/collected data and labeled. ● Key attrition-related features are selected. ● Classification models are trained and evaluated. ● Attrition risk scores are generated. ● Retention strategies reduce hiring and training costs. Customer Segmentation (Marketing): <ul style="list-style-type: none"> ● Customer demographic and behavioral data is collected. ● Clustering algorithms group similar customers. ● Segments are interpreted based on business value. 	15T

	<ul style="list-style-type: none"> ● Targeted marketing strategies are deployed. ● Marketing ROI is continuously optimized. 	
III	<p>Forecasting and Deep Learning as Automation: Future prediction and intelligent automation; Time Series Concepts and Data Preparation; Trend, Seasonality, and Noise Forecasting Models, AI Applications in Healthcare.</p> <p>Product Price Forecasting (Sales Department)</p> <ul style="list-style-type: none"> ● Historical product price data. ● Patterns are analyzed. ● Forecasting models learn historical behavior. ● Future prices are predicted with confidence. ● Pricing and inventory decisions are improved. <p>Disease Detection</p> <ul style="list-style-type: none"> ● Medical images data are used/collected. ● CNNs automatically learn discriminative features. ● Disease patterns are detected and classified. ● Predictions assist clinicians in diagnosis. ● Diagnostic accuracy and efficiency are enhanced. 	15T
IV	<p>NLP and Computer Vision as Intelligent Perception: Natural Language Processing Fundamentals, Text Cleaning, Tokenization, and Representation, Sentiment Analysis Techniques, Social Media and Opinion Mining, Model Deployment and Performance Monitoring.</p> <p>Sentiment Analysis of Customer Reviews (Public Relations)</p> <ul style="list-style-type: none"> ● Social media and review data is collected. ● Text data is preprocessed and vectorized. ● NLP models classify sentiment polarity. ● Sentiment trends are monitored over time. ● Brand reputation and customer engagement improve. <p>Defect Detection in Manufacturing (Production/Maintenance)</p> <ul style="list-style-type: none"> ● Product images are captured on production lines. ● Image preprocessing enhances defect visibility. ● Vision models detect and classify defects. ● Defect localization supports corrective action. ● Quality control and operational efficiency increase. 	15T
Pedagogy :	Lectures/tutorials/ assignments/ mini-projects/ PPT presentations/ case studies/ class discussions.	
References/ Reading :	1. Sharda, R., Delen, D., & Turban, E. (2020). <i>Business intelligence, analytics, and data science: A managerial perspective</i> (4th ed.).	

	<p>Pearson Education.</p> <p>2. Davenport, T. H., & Harris, J. G. (2017). <i>Competing on analytics: The new science of winning</i> (2nd ed.). Harvard Business School Press.</p> <p>3. Marr, B. (2016). <i>Data strategy: How to profit from a world of big data, analytics and the Internet of Things</i>. Kogan Page.</p>
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. Identify and interpret core thematic ideas in data science across multiple domains.</p> <p>CO 2. Explain the end-to-end working of data science applications using real-world datasets.</p> <p>CO 3. Apply appropriate thematic approaches to structured and unstructured data problems.</p> <p>CO 4. Apply data science solutions aligned with organizational and societal needs.</p>



Minor VET Courses

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-322
Title of the Course : Big Data
Number of Credits : 4(3T+1P)
Effective from AY : 2024-25

Pre-requisites for the Course :	IDS-142 Programming in Python	
Course Objectives :	The course aims to provide students with a comprehensive understanding of Big Data, including the challenges involved in its storage and analysis, while offering hands-on experience with big data platforms and tools.	
Units	Content	No. of hours 75 (45T+30P)
I	<p>Introduction to Big Data: Big Data and the V's of Big Data, Big Data Analytics, Challenges faced in storage, querying and analysis of big data; need for big data frameworks</p> <p>Hadoop Framework: Apache Hadoop Basics: Introduction, Hadoop 1 v/s Hadoop 2, Use cases and anti-patterns, Main components of Hadoop, Hadoop Ecosystem</p> <p>Hadoop Distributed Filesystem (HDFS): Architecture, daemons and their roles, data blocks, replication policy, Handling node & disk failures, Namenode startup operation, checkpointing process, file read & write process, Error handling in read/write, NameNode resilience</p> <p>Mapreduce: Introduction to the MapReduce framework, Map, Shuffle-Sort and Reduce Phases, IData Flow in a Map Reduce Job, Daemons, Partitions, Combiner functions.</p> <p>Apache YARN: YARN applications, YARN daemons, Anatomy of a YARN application run</p>	15T
II	<p>MapReduce Programming</p> <p>Mapreduce Programming in Java: Hadoop Data Types, Input & Output formats, Overview of a Map Reduce program</p> <p>Advanced Mapreduce: Chaining jobs, Joining data, Secondary sorting.</p>	15T

<p>III</p>	<p>Hadoop Ecosystem:</p> <p>Apache PIG:Introduction, Execution Modes, Data Model (Relations, Bags, Tuples, Fields), Pig Latin Basics.</p> <p>Apache Hive: Introduction, Hive Architecture, Metastore, Schema on Read, Hive Tables (Managed and External), Partitions, Buckets.</p> <p>Apache HBase: Introduction, Data Model, Architecture, Metadata, API (Get, Put, Scan)</p> <p>Apache Spark:</p> <p>Apache Spark Basics: Introduction, Use Cases and antipatterns, Spark Components, Spark Program Flow, Resilient Distributed Dataset (RDD), Actions & Transformations on Basic and Pair RDD.</p> <p>SparkQL: Spark SQL vs. Traditional SQL Databases, DataFrames and Datasets, Spark SQL Functions and Queries</p>	<p>15T</p>
<p>IV</p>	<p>List of suggested assignments:</p> <ol style="list-style-type: none"> 1. Install Hadoop in the pseudo-distributed mode. 2. Using HDFS commands, implement file management tasks in Hadoop 3. Write Word Count MapReduce program 4. Write Map Reduce programs to <ol style="list-style-type: none"> a. Find the average rating on a movie data set b. Find the number of times each user of age > 40 has rated a movie . c. Implement Matrix Multiplication d. Mine weather data. 5. Install Pig and write a Pig Latin script to sort, group, join, project, and filter your data. 6. Install Hive, then use Hive to create, alter, and drop databases, tables, views, functions, and indexes. Using HiveQL, sort, group, join, project, and filter your data. 7. Install pySpark (can be done using Docker container) and create RDDs, perform actions and transformations on the RDDs 8. Using pySpark, load a text file as a RDD and perform a word count. 9. Convert a JSON file into a DataFrame. Run SQL queries to sort, group, join, project, and filter your data. 	<p>30P</p>

	<p>Mini project</p> <p>Solve some real-life big data problems.</p> <ul style="list-style-type: none"> ● Traffic control using big data ● Medical insurance fraud detection ● Recommendation system ● Anomaly detection in cloud servers ● Tourist behavior analysis ● Web server log analysis 	
Pedagogy:	Lectures/Assignments/Flipped classroom/Lab assignments/mini project	
References/ Reading :	<ol style="list-style-type: none"> 1. White, T. (2015). Hadoop: The definitive guide (4th ed.). O'Reilly Media. 2. Perrin, J.-G. (2020). Spark in action (2nd ed.). Manning Publications. 3. Lam, C. (2011). Hadoop in action. Manning Publications. 4. Holmes, A. (2012). Hadoop in practice. Manning Publications. 5. DeRoos, D., Zikopoulos, P., Brown, B., Coss, R., & Melnyk, R. B. (2014). Hadoop for dummies. John Wiley & Sons. 6. Overview. (2017). Apache.org. https://pig.apache.org/docs/latest/index.html 7. RDD Programming Guide - Spark 3.0.0 Documentation. (n.d.). Spark.apache.org. https://spark.apache.org/docs/latest/rdd-programming-guide.html 	
Course Outcomes :	<p>On completion of the course, students will be able to:</p> <p>CO 1. Explain the fundamentals of Big Data, including the core principles, concepts, and technologies that underpin big data analytics.</p> <p>CO 2. Describe the principles, concepts, and technologies that form the basis of big data analytics platforms.</p> <p>CO 3. Apply distributed computing frameworks such as Apache Hadoop and Apache Spark to process and transform large-scale datasets.</p> <p>CO 4. Analyze case studies and real-world scenarios to evaluate the practical applications of big data analytics.</p>	

Name of the Programme : Master of Science (Integrated) in Data Science
Course Code : IDS-323
Title of the Course : Usability and Quality Assurance
Number of Credits : 4 (3T+1P)
Effective from AY : 2024-25

Pre-requisites for the Course :	IDS-111 Problem Solving and Program Design IDS-142 Programming in Python IDS-204 Fundamentals of Computing Systems Design IDS-321 Software Engineering	
Course Objectives :	The course is aimed at providing learners with the necessary exposure to the job responsibilities of Software Quality Assurance (QA) engineers & User Experience (UX) designers in the IT industry	
Units	Content	No. of hours 75 (45T+30P)
I	Fundamentals of Software Testing: Role of QA in SDLC & Agile/DevOps. Manual vs Automated Testing. Test levels and test types. Test case design techniques Developer-Centric Testing Concepts : Unit testing principles. Test-Driven Development (TDD). MVC testing concepts. API testing fundamentals.	15T
II	Test Automation Architecture : Test automation challenges. Design patterns in test automation. Page Object Model. Business Layer Page Object. Overview of test automation frameworks. BDD & Business-Driven Testing concepts Non-Functional Testing Concepts: Performance, load, and stress testing. Log analysis and monitoring basics. Human-Centered Design : Gulf of execution & evaluation. Human error vs bad design. Double-Diamond Model. Universal design principles.	15T
III	Visual Design Foundations : Visual design elements. Line, color, shape, form, space, texture, value. Extended elements. Typography, movement. Visual design principles. Balance, harmony, dominance, scale. UX Process & Tools Overview : Wireframes, mockups, prototypes. Interaction & animation basics. Raster vs vector	15T

	graphics. UX portfolio principles	
IV	<p>List of Suggested Lab Assignments</p> <p>Lab assignments demonstrating application of following concepts :-</p> <ul style="list-style-type: none"> ● Manual test plan & test cases for a web app ● Automated test scripts for key workflows (Selenium) ● API test suite with assertions (Postman, SOAP-UI, JUnit, NUnit, MVC component testing) ● Test Frameworks like TestNG, Cucumber ● BDD test suite using tools like Gherkin ● Log analysis ● Performance testing using tools like JMeter ● Mobile app testing using tools like Appium ● UX case study analysis (personas, user journeys) ● Wireframe for a real-world app (paper prototypes, Figma, Adobe XD) ● High-fidelity UI mockup (typography, colors, layout, interaction) ● Graphic asset design (vector & raster graphics) 	30P
Pedagogy :	Lectures, Tutorials. Assignments, Seminars, Presentations or a combination of some of these can also be used. ICT mode should be preferred. Sessions should be interactive in nature to enable peer group learning.	
References/ Reading :	<ol style="list-style-type: none"> 1. Graham, D., Black, R., & van Veenendaal, E. (2020). Foundations of software testing: ISTQB certification (4th ed., updated for ISTQB Foundation Syllabus 2018). Cengage Learning. 2. Norman, D. A. (2013). The design of everyday things (Rev. and expanded ed.). MIT Press. 3. Gatto, J. A., Porter, A. W., & Selleck, J. (2011). Exploring visual design: The elements and principles (4th ed.). Davis Publications. 	
Course Outcomes :	<p>On completion of the course, students will be able to :</p> <p>CO 1. understand software testing and quality assurance as a fundamental component of software life cycle</p> <p>CO 2. efficiently perform quality assurance activities using modern software tools</p> <p>CO 3. prepare test plans and schedules for a quality assurance project</p> <p>CO 4. perform quality assurance activities using modern software tools</p>	