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)

GU/Acad –PG/BoS -NEP/2025-26/230 Date: 07.07.2025

CIRCULAR

The Academic Council & Executive Council of the University has approved Ordinance OA-35A relating to PG Programmes offered at the University campus and its affiliated Colleges based on UGC 'Curriculum and Credit Framework for Postgraduate Programmes'. Accordingly, the University has proposed introduction of Ordinance OA-35A from the Academic year 2025-2026 onwards.

The Programme structure and syllabus of Semester I and II of the **Master of Science in Data Science** Programme approved by the Academic Council in its meeting held on 13th & 14th June 2025 is attached.

The Dean & Vice-Dean (Academic) 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 VYAS LAWANDE
Digitally signed by ASHWIN VYAS LAWANDE
Date: 2025.07.07 15:15:42 +05'30'

(Ashwin V. Lawande)
Deputy Registrar – Academic

To,
1. The Dean, Goa Business School, Goa University.
2. The Vice-Dean (Academic), Goa Business School, Goa University.

Copy to:
1. Chairperson, BoS in Data Science and Artificial Intelligence, Goa University.
2. Programme Director, M.Sc. Data Science, Goa University.
3. Controller of Examinations, Goa University.
4. Assistant Registrar Examinations (PG), Goa University.
5. Director, Directorate of Internal Quality Assurance, Goa University for uploading the Syllabus on the University website.

GOA UNIVERSITY

MASTER OF SCIENC IN DATA SCIENCE

(Effective from the Academic Year 2025-2026)

ABOUT THE PROGRAMME

The M.Sc. Data Science programme is a two-year postgraduate programme designed for undergraduates in Computer Science, IT, Data Science, Artificial Intelligence, related Science disciplines, and Engineering streams. It provides a strong foundation in statistics, mathematics, and computer science, equipping students with practical skills in data analysis, machine learning, big data technologies, and data-driven decision-making.

The programme emphasizes interdisciplinary learning, hands-on experience with modern tools, and the development of critical thinking, ethical awareness, and analytical reasoning. Graduates will be prepared for careers in industry, research, or academia, with the capacity to drive innovation, entrepreneurship, and research-led solutions in the data science domain.

OBJECTIVES OF THE PROGRAMME

1. To provide a strong understanding of core concepts in statistics, mathematics, and computer science that serve as a foundation for data science practices.
2. To train students with practical skills in using modern tools, programming environments, and big data technologies for the end-to-end lifecycle of data, from collection and processing to analysis, interpretation, and visualization in real-world applications.
3. To cultivate critical thinking and analytical reasoning necessary for designing data-driven solutions to diverse real-world challenges.
4. To promote a research-oriented mindset through systematic problem formulation, experimental design, and analytical evaluation.

Approved by: Academic Council on 13th & 14th June 2025

PROGRAMME SPECIFIC OUTCOMES (PSO)	
PSO 1.	Demonstrate fundamental knowledge of statistics, mathematics, and computer science concepts essential for data science.
PSO 2.	Apply data analysis tools and software to manage, process, and analyze data.
PSO 3.	Cultivate critical thinking and analytical reasoning skills to design data science-enabled solutions.
PSO 4.	Develop proficiency in implementing machine learning algorithms and models to address real-world challenges across diverse domains.
PSO 5.	Develop expertise in visualizing complex datasets and effectively communicating data-driven insights.
PSO 6.	Foster a research-oriented mindset, enabling the formulation of research problems, experimentation, and analysis to address emerging research challenges in data science.
PSO 7.	Apply ethical principles, ensuring their data science practices respect privacy and align with societal values and legal standards.
PSO 8.	Apply data science expertise to contribute to addressing societal issues

PROGRAMME STRUCTURE

Master of Science in Data Science

Effective from Academic Year 2025-26

SEMESTER I				
Discipline Specific Core (DSC) Courses (16 credits)				
Sr. No.	**Course Code**	**Title of the Course**	**Credits**	**Level**
1	CSD-5000	Principles of Data Science	4T	400
2	CSD-5001	Mathematical Foundations for Data Science	4T	400
3	CSD-5002	Problem Solving and Programming Fundamentals	2T	400
4	CSD-5003	Data Science Lab	3P	400
5	CSD-5004	Mathematical Foundations for Data Science Lab	3P	400
Total Credits for DSC Courses in Semester I			**16**	
Discipline Specific Elective (DSE) Course (4 credits)				
Sr. No.	**Course Code**	**Title of the Course**	**Credits**	**Level**
1	CSD-5201	Data Management	4T	400
2	CSD-5202	Algorithms and Computational Complexity	4T	400
Total Credits for DSE Courses in Semester I			**4**	
Total Credits in Semester I			**20**	

SEMESTER II				
Discipline Specific Core (DSC) Courses				
Sr. No.	**Course Code**	**Title of the Course**	**Credits**	**Level**
1	CSD-5005	Machine Learning Techniques	4T	500
2	CSD-5006	Data Security, Privacy, and Ethics	2T	500
3	CSD-5007	Data Modeling and Visualization	2T	500
4	CSD-5008	Optimization Techniques	4T	500
5	CSD-5009	Machine Learning Techniques Lab	2P	500
6	CSD-5010	Data Modeling and Visualization Lab	2P	500
Total Credits for DSC Courses in Semester II			**16**	
Discipline Specific Elective (DSE) Courses (4 credits)				
Sr. No.	**Course Code**	**Title of the Course**	**Credits**	**Level**
1	CSD-5203	Domain-Specific Predictive Analytics	4T	400
2	CSD-5204	Web Data Analytics	4T	400
Total Credits for DSE Courses in Semester II			**4**	
Total Credits in Semester II			**20**	

Blooms Taxonomy Cognitive Levels	
Cognitive Level	**Notations**
K1	Remembering
K2	Understanding
K3	Applying
K4	Analyzing
K5	Evaluating
K6	Create

SEMESTER I

Discipline Specific Core Courses

Title of the Course	Principles of Data Science
Course Code	CSD-5000
Number of Credits	4
Theory/Practical	Theory
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	This course introduces the foundational principles, processes, and tools of data science, covering the lifecycle from data collection and preprocessing to analysis, modeling, and evaluation.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand the data science lifecycle, key concepts, and the interdisciplinary role of a data scientist.	PSO1, PSO2
	CO 2. Apply data preprocessing techniques for cleaning, transforming, and preparing structured and unstructured data.	PSO2, PSO3

	CO 3. Analyze data using descriptive, diagnostic, and predictive techniques, and build simple models using regression and classification.		PSO2, PSO3, PSO4	
	CO 4. Evaluate model performance using appropriate metrics and understand ethical, privacy, and security considerations in data science applications.		PSO3, PSO7, PSO8	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	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	**15**	CO1	K1, K2, K3
Module 2:	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, and dimensionality reduction.	**15**	CO2	K1, K2, K3
Module 3:	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.	**15**	CO3	K3, K4
Module 4:	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.	**15**	CO4	K1, K4, K5
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom			
Texts:	1. Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press. 2. Blum, A., Hopcroft, J., & Kannan, R. (2020). Foundations of data science. Cambridge University Press.			

References/ Readings:	1. Pierson, L. (2021). Data science for dummies. John Wiley & Sons. 2. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. O'Reilly Media
Web Resources:	OpenStax. (n.d.). Principles of data science. OpenStax. Retrieved May 9, 2025, from https://openstax.org/details/books/principles-data-science

Title of the Course	Mathematical Foundations for Data Science
Course Code	CSD-5001
Number of Credits	4
Theory/Practical	Theory
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	To build a strong mathematical foundational knowledge in probability, statistics, calculus, and linear algebra essential for modeling, analyzing, and solving data-driven problems in data science.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Apply fundamental principles of probability to understand the uncertainty in data-driven problems.	PSO1, PSO3
	CO 2. To apply statistical inference techniques to analyze and interpret data from both large and small samples	PSO1, PSO2, PSO3
	CO 3. Understand the use of calculus in the mathematical modeling of machine learning algorithms.	PSO1, PSO4
	CO 4. Apply linear algebra concepts to efficiently represent and process data.	PSO1, PSO2, PSO4

Content:		No of hours	Mapped to CO	Cognitive Level
Module 1:	**Probability and Statistics** Basic concepts of probability, conditional probability, law of total probability, independence of events, Bayes' theorem, random variables (discrete and continuous), expectation, moments, moment generating functions, commonly used probability distributions, joint and conditional distributions, transformation of random variables, covariance and correlation. Sampling techniques, sampling distributions.	**15**	CO1	K2, K2, K4
Module 2:	**Advanced Statistics** Parameter estimation, Methods of point estimation, Methods of Interval estimation, hypothesis testing, Large sample tests, Small sample tests, Non-parametric tests.	**15**	CO2	K3, K4, K5
Module 3:	**Calculus** Functions of a single variable, limit, continuity, differentiability, Mean value theorems, indeterminate forms, L'Hospital's rule, Maxima and minima, Product and chain rule, Taylor's series, infinite series summation/integration concepts, Fundamental and mean value-theorems of integral calculus, evaluation of definite and improper integrals, Beta and gamma functions, Functions of multiple variables, limit, continuity, partial derivatives, Basics of ordinary and partial differential equations	**15**	CO3	K2, K3
Module 4:	**Linear Algebra** Representation of vectors and matrices, linear dependence and independence, vector spaces and subspaces (definition, examples, and basis), linear transformations, range and null space, special types of matrices, eigenvalues and eigenvectors, diagonalization, singular value decomposition (SVD), least squares and minimum norm solutions, applications to data analysis.	**15**	CO4	K2, K3, K4
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom			
Texts:	1. Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2024). Probability and statistical inference 10th Edition, Pearson. 2. Srivastava, M. K., & Srivastava, N. (2009). Statistical Inference: Testing of Hypotheses. PHI Learning Pvt. Ltd..			

	3. Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for machine learning (1st ed.). Cambridge University Press. 4. Nicholson, W. K. (2023). Linear algebra with applications. Boston: PWS Publishing Company.
References/ Readings:	1. Ross, S. M. (2020). A first course in probability. Pearson. 2. Casella, G., & Berger, R. (2024). Statistical inference. CRC Press. 3. Strang, G. (2022). Introduction to linear algebra. Wellesley-Cambridge Press.
Web Resources:	MIT OpenCourseWare. (n.d.). Linear algebra [Video series]. Massachusetts Institute of Technology. Retrieved May 10, 2025, from https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/video_galleries/video-lectures/

Title of the Course	Problem Solving and Programming Fundamentals
Course Code	CSD-5002
Number of Credits	2
Theory/Practical	Theory
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	To equip learners with foundational concepts and programming skills necessary for designing and implementing solutions to data-intensive problems using appropriate tools, techniques, and data structures in data science.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand the fundamental concepts of problem-solving and programming constructs such as algorithms, flowcharts, and pseudocode.	PSO1
	CO 2. Apply structured problem-solving techniques to develop algorithmic solutions using appropriate control structures.	PSO1, PSO2
	CO 3. Design and represent solutions to computational problems using flowcharts, pseudocode, and basic data structures.	PSO1, PSO3
	CO 4. Analyze problem scenarios to design, test, and refine modular and structured solutions	PSO1, PSO3, PSO6

	using programming tools.			
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	**Problem Solving Basics:** Problem-solving process; Identifying solution types; Approaches for solving problems using computers. Constants, variables, data types, and functions. Operators, expressions, and equations used in problem-solving. **Solution Planning and Design:** Communicating with the computer; Tools for planning solutions (flowcharts, pseudocode). Introduction to the Software Development Life Cycle (SDLC). **Techniques of Problem Solving:** Sequential Problem Solving: Using flowcharts, algorithmic instructions, and pseudocode. Decision Problem Solving: Logic types, decision tables. Iterative Problem Solving: Loops, recursion, incrementing & accumulating, types of loops.	**15**	CO1 CO2 CO3	K2, K3
Module 2:	**Composite Data Structures:** Introduction to lists, arrays, and maps (one-dimensional, two-dimensional arrays), Basic concepts of linked lists, graphs, and trees. Database and File Handling: Introduction to DBMS and file processing concepts. **Learning by Tools:** Using tools like PictoBlox, Scratch, TurtleArt, and programming languages like Kojo and Racket (DrRacket); Hands-on assignments: Create geometric shapes, solve arithmetic problems like factorials, prime numbers, etc.	**15**	CO3 CO4	K3, K4
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom			
Texts:	Felleisen, M., Findler, R. B., Flatt, M., & Krishnamurthi, S. (2018). How to design programs: an introduction to			

	programming and computing. MIT Press.
References/ Readings:	1. Abelson, H., & Sussman, G. J. (1996). Structure and interpretation of computer programs (p. 688). The MIT Press. 2. Maureen Sprankle, Jim Hubbard (2013). Problem Solving and Programming Concepts. Pearson Education India. Latest Edition. 3. Kuppuswamy, S., MaIliga, S., Kanimozhi Selvi, C. S., & KousaIya, K. (2019). Problem Solving and Programming. Tata McGraw Hill.
Web Resources:	Felleisen, M., Findler, R. B., Flatt, M., & Krishnamurthi, S. (2001). How to design programs. Retrieved May 9, 2025, from https://htdp.org/

Title of the Course	Data Science Lab
Course Code	CSD-5003
Number of Credits	3
Theory/Practical	Practical
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	To provide hands-on experience in Python programming, data analysis, and database management, empowering students to use Python libraries, build applications, and perform SQL queries for solving real-world data science problems.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Apply Python programming concepts to solve real-world problems, including the use of control structures, functions, recursion, and object-oriented programming.	PSO1, PSO2, PSO3
	CO 2. Analyze and manipulate datasets using Python libraries to perform data cleaning, transformation, and visualization.	PSO2, PSO3, PSO5
	CO 3. Design and develop interactive applications using Streamlit, Flask, and Django to present data-driven insights.	PSO2, PSO3, PSO5, PSO6

	CO 4. Evaluate and perform SQL queries to manage, retrieve, and analyze data from relational databases.		PSO2, PSO3	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	Programming in Python 1. Set up a UNIX/Python environment to run and manage Python programs. 2. Implement code collaboration workflows using Git for version control. 3. Practice decision-making, branching, and looping constructs using control structures. 4. Work with Python data types including lists, sets, tuples, dictionaries, and strings. 5. Implement user-defined functions and recursive solutions for computational problems. 6. Create and manipulate user-defined data types and perform file handling operations. 7. Perform CRUD operations using text files for data storage and retrieval. 8. Create a custom data type using classes and objects in Python. 9. Implement unit testing to verify the correctness and robustness of Python code.	**30**	CO1	K3, K4
Module 2:	Python Libraries, Packages, Frameworks 1. Use Pandas library to handle, clean, and manipulate structured data. 2. Work with Python packages for loading and preprocessing real-world datasets. 3. Apply time-series analysis techniques to explore temporal patterns in data. 4. Visualize data using various plots, charts, and graphs to uncover insights. 5. Develop interactive data-driven applications using Streamlit. 6. Build and deploy simple web applications using the Flask framework. 7. Create scalable and structured web applications using the Django framework.	**30**	CO2 CO3	K4. K5
Module 3:	Querying Data (SQL) 1. Set up a relational database and create, modify tables with appropriate constraints. 2. Query and explore metadata to understand the structure and schema of a database.	**30**	CO4	K4. K5

	3. Perform CRUD operations to insert, update, delete, and retrieve data from tables. 4. Retrieve specific data using operators, wildcards, sorting, and built-in SQL functions. 5. Use joins, aggregate functions, grouping, and sub-queries to analyze and summarize data.			
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom/Mini-project			
Texts:	1. Sweigart, A. (2019). Automate the boring stuff with Python: Practical programming for total beginners (2nd ed.). No Starch Press. 2. Beaulieu, A. (2009). Learning SQL. O'Reilly Media. Latest Edition.			
References/ Readings:	1. McKinney, W. (2017). Python for data analysis: Data wrangling with pandas, numpy, and ipython (2nd ed.). O'Reilly Media. 2. Garcia-Molina, H. (2008). Database systems: the complete book. Pearson Education India.			
Web Resources:	1. Data.World. (n.d.). Introduction to SQL concepts. Data.World. Retrieved May 9, 2025, from https://docs.data.world/documentation/sql/concepts/basic/intro.html 2. TutorialsPoint. (n.d.). SQL tutorial. TutorialsPoint. Retrieved May 9, 2025, from https://www.tutorialspoint.com/sql/index.htm 3. SQLZoo. (n.d.). SQL tutorial. SQLZoo. Retrieved May 9, 2025, from https://sqlzoo.net/wiki/SQL_Tutorial 4. Python Software Foundation. (n.d.). Python documentation. Python.org. Retrieved May 9, 2025, from https://docs.python.org/3/ 5. Streamlit Inc. (n.d.). Streamlit documentation. Streamlit. Retrieved May 9, 2025, from https://docs.streamlit.io/ 6. Pallets Projects. (n.d.). Flask tutorial. Flask Documentation. Retrieved May 9, 2025, from https://flask.palletsprojects.com/en/stable/tutorial/ 7. Django Software Foundation. (n.d.). Django documentation. Django. Retrieved May 9, 2025, from https://docs.djangoproject.com/			

Title of the Course	Mathematical Foundations for Data Science Lab
Course Code	CSD-5004
Number of Credits	3
Theory/Practical	Practical
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil			
Course Objectives:	To provide hands-on experience in implementing foundational mathematical and statistical concepts through computational tools for data analysis, visualization, inference, and dimensionality reduction on real-world datasets.			
Course Outcomes:			**Mapped to PSO**	
	CO 1. Simulate and visualize probabilistic models and distributions using computational tools.		PSO1	
	CO 2. Apply sampling techniques on real-world datasets and interpret the results.		PSO1, PSO2	
	CO 3. Perform hypothesis testing and draw statistical inferences on real world data.		PSO1, PSO2	
	CO 4. Implement dimensionality reduction techniques on real world dataset		PSO1, PSO2	
Content:	Use Python libraries such as NumPy, SciPy, Pandas, Matplotlib, Seaborn to perform the following experiments:	**No of hours**	**Mapped to CO**	**Cognitive Level**

Module 1:	1. Simulate basic probability experiments to apply conditional probability and total probability. 2. Visualize probability distributions and explore their properties. 3. Implement various sampling techniques on real world data. 4. Using the chi-square test, check goodness-of-fit and test of independence.	**30**	CO1, CO2	K2, K3
Module 2:	1. Perform Z-test and t-test for single population mean. 2. Implement One-sample and two-sample t-tests 3. Perform F-test and Levene's test to compare variances	**30**	CO3	K2, K3
Module 3:	1. Represent a real world dataset as vectors or matrices. Perform linear transformations and visualize them. 2. Implement Principal Component Analysis and reduce the dimensionality of a real world dataset 3. Implement Singular Value Decompositions and reduce the dimensionality of a real world dataset.	**30**	CO4	K3
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom/Mini-project			
Texts:	1. Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2024). Probability and statistical inference 10th Edition, Pearson. 2. Srivastava, M. K., & Srivastava, N. (2009). Statistical Inference: Testing of Hypotheses. PHI Learning Pvt. Ltd.. 3. Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for machine learning (1st ed.). Cambridge University Press. 4. Nicholson, W. K. (2023). Linear algebra with applications. Boston: PWS Publishing Company.			
References/ Readings:	1. Downey, A. (2014). Think stats: Exploratory data analysis. 2nd Edition " O'Reilly Media, Inc.". 2. Wes, M. (2017). Python for data analysis. 2nd Edition " O'Reilly Media, Inc.".			
Web Resources:	1. Statquest. (n.d.). Statistics fundamentals [YouTube playlist]. YouTube. Retrieved May 10, 2025, from			

	https://www.youtube.com/playlist?list=PLblh5JKOoLUK0FLuzwntyYI10UQFUhsY9 2. 3Blue1Brown. (n.d.). Essence of linear algebra [YouTube playlist]. YouTube. Retrieved May 10, 2025, from https://www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab 3. MIT OpenCourseWare. (n.d.). Linear algebra [Video series]. Massachusetts Institute of Technology. Retrieved May 10, 2025, from https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/video_galleries/video-lectures/

Discipline Specific Elective Courses

Title of the Course	Data Management
Course Code	CSD-5201
Number of Credits	4
Theory/Practical	Theory
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	This course aims to equip students with the knowledge of relational and non-relational database systems, data modeling, data retrieval techniques, and modern data storage paradigms, including NoSQL and in-memory databases.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand the structure, types, and lifecycle of data in the context of data science workflows.	PSO1, PSO2
	CO 2. Analyze and apply database design principles using entity-relationship modeling and relational schema translation.	PSO1, PSO3
	CO 3. Apply normalization techniques to design efficient and consistent relational databases.	PSO1, PSO3

	CO 4. Construct and optimize SQL queries for data retrieval, aggregation, and manipulation.		PSO2, PSO3	
	CO 5. Explain the concepts and applications of transactions, indexing, and query optimization in relational databases.		PSO1, PSO3	
	CO 6. Compare relational databases with NoSQL and in-memory databases, identifying their appropriate use cases in data science applications.		PSO1, PSO2, PSO3	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	Foundations of Data Management: Data and Information, Hierarchy of Data (Bit, Byte, Field, Record, File, Database), Data Types and Representations (Structured, Semi-structured, Unstructured) Data Lifecycle and Workflow in Data Science Data Handling Tools in Data Science, Data Storage and Retrieval, Data Security, Access Control, and Privacy Considerations	**15**	CO1	K1, K2
Module 2:	Database Design and Relational Modeling: Database Concepts and Terminology, Data Models, ER Modeling and ER Diagrams, Translating ER Models to Relational Schemas Relational Model: Tables, Keys, Integrity Constraints Functional Dependencies, Normalization: 1NF, 2NF, 3NF	**15**	CO2, CO3	K2, K3, K4
Module 3:	Data Retrieval and Query Processing: SQL Basics: SELECT, WHERE, ORDER BY, GROUP BY Advanced Queries: JOINs (INNER, OUTER, SELF), Subqueries Aggregate Functions, GROUP BY and HAVING Clauses Views, Indexes, and Stored Procedures Transactions and ACID Properties; Query Optimization Basics	**15**	CO4, CO5	K2, K3, K4
Module 4:	NoSQL and In-Memory Databases: NoSQL Data Models, NoSQL vs SQL, Basics of Data Replication and Sharding, CAP	**15**	CO6	K2, K4, K5

	Theorem and BASE vs ACID In-Memory Databases: Concept, Use Cases, and Performance			
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom			
Texts:	1. Silberschatz, A., Korth, H. F., & Sudarshan, S. (2020). Database system concepts (7th ed.). McGraw-Hill Education. 2. Kleppmann, M. (2017). Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. O'Reilly Media.			
References/ Readings:	Elmasri, R., & Navathe, S. B. (2017). Fundamentals of database systems (7th ed.). Pearson Education.			
Web Resources:	MIT OpenCourseWare - Database Systems (2018). Retrieved on May 13, 2025, from https://ocw.mit.edu/courses/6-830-database-systems-fall-2010/			

Title of the Course	Algorithms and Computational Complexity
Course Code	CSD-5202
Number of Credits	4
Theory/Practical	Theory
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	To introduce the fundamental principles of algorithmic problem solving, including complexity analysis, searching, sorting, and algorithmic strategies relevant to data science.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand fundamental algorithmic concepts, time and space complexity, and recursive structures.	PSO1, PSO2
	CO 2. Apply searching and sorting algorithms to solve basic data-centric problems.	PSO2, PSO3
	CO 3. Utilize appropriate data structures and graph algorithms to address computation-heavy tasks.	PSO3
	CO 4. Analyze and implement algorithmic design strategies like greedy, divide-and-conquer, and dynamic programming.	PSO2, PSO3

	CO 5. Evaluate the efficiency and scalability of string matching algorithms, hashing techniques, and heap-based algorithms for real-world data science applications.		PSO4, PSO5	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	**Foundations of Algorithms and Complexity:** Introduction to Algorithms: Definition, characteristics, and importance in data science Algorithm Representation: Pseudocode, flowcharts Time and Space Complexity: Big O, Big Theta, Big Omega Recursion and Recursive Algorithms Searching Algorithms: Linear Search, Binary Search Sorting Algorithms: Bubble, Selection, Insertion, Merge Sort, Quick Sort Stability and Efficiency of Sorting Algorithms	**15**	CO1, CO2	K2, K3
Module 2:	**Advanced Data Structures and Graph Algorithms:** Linear Data Structures: Arrays, Stacks, Queues. Non-Linear Data Structures: Trees, Graphs Graph Representations: Adjacency list/matrix Graph Traversal Algorithms: Breadth-First Search (BFS), Depth-First Search (DFS)	**15**	CO3	K3, K4
Module 3:	**Algorithm Design Techniques** Greedy Algorithms: Principle, examples (knapsack problem) Divide and Conquer: Concept and examples (merge sort, binary search) Dynamic Programming: Overlapping subproblems and optimal substructure (e.g., knapsack problem) Backtracking and Branch and Bound (e.g., N-Queens, TSP introduction)	**15**	CO4	K3, K4

Module 4:	**Algorithm Applications in Data Science** String Matching Algorithms: KMP Hashing Techniques and Hash Functions Introduction to Heaps and Priority Queues Case Studies: Algorithmic Applications in Data Science	**15**	CO5	K3, K4, K5
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom			
Texts:	Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms* (3rd ed.). MIT Press.			
References/ Readings:	1. Skiena, S. S. (2020). The algorithm design manual (3rd ed.). Springer. 2. Goodrich, M. T., Tamassia, R., & Goldwasser, M. H. (2013). *Data structures and algorithms in Python*. Wiley. 3. Kleinberg, J., & Tardos, É. (2006). *Algorithm design*. Pearson Education.			
Web Resources:	1. MIT OpenCourseWare- Introduction to Algorithms. Retrieved May 13, 2025, from https://ocw.mit.edu/courses/6-006-introduction-to-algorithms-spring-2020/			

SEMESTER II

Discipline Specific Core Courses

Title of the Course	Machine Learning Techniques
Course Code	CSD-5005
Number of Credits	4
Theory/Practical	Theory
Level	500
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	CSD-5001 and CSD-5002	
Course Objectives:	To equip students to understand and apply key machine learning techniques—such as classification, regression, clustering, neural networks, and reinforcement learning—to analyze data and solve real-world problems.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand the fundamental concepts, types, and goals of machine learning and explain key components like hypothesis and version space.	PSO1, PSO2
	CO 2. Apply supervised learning algorithms such as decision trees, random forests, linear and logistic regression, and support vector machines to classification and regression tasks.	PSO2, PSO3

	CO 3. Evaluate machine learning models using performance metrics, cross-validation, and address issues like overfitting and underfitting.		PSO2, PSO4	
	CO 4. Implement ensemble learning techniques such as bagging and boosting to improve model accuracy and robustness.		PSO3, PSO4	
	CO 5. Analyze and solve unsupervised learning problems using clustering algorithms and dimensionality reduction with PCA and Gaussian Mixture Models.		PSO2, PSO4	
	CO 6. Understand and apply neural network architectures and reinforcement learning methods for sequential decision-making problems.		PSO4, PSO6	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	**Introduction to Machine Learning** Definition of Learning Systems, Goals and Applications of Machine Learning, Types of Machine Learning: Supervised, Unsupervised, and Reinforcement Learning, Hypothesis Space and Version Space Concept Learning: Concept Learning Task, Notation, Inductive Learning Hypothesis, Concept Learning as Search, Version Space and Candidate Elimination Algorithm, Decision Trees, Random Forest Linear Regression: Simple Linear Regression, Multiple Linear Regression, Cost Function, Gradient Descent, Evaluating Models using Metrics: Mean Squared Error (MSE), R-Squared	**15**	CO1, CO2	K2, K3
Module 2:	**Supervised Learning:** Logistic Regression: Introduction and Application of Logistic Regression, Sigmoid Function and Decision Boundary, Cost Function and Optimization (Gradient Descent) Support Vector Machines (SVM): Introduction, Linear SVM: Maximizing the Margin, Kernel Methods: Non-linear SVM, Soft Margin SVM and Regularization Ensemble Methods: Bagging: Bootstrap Aggregating (Random Forest), Boosting: AdaBoost, Gradient Boosting, Evaluation of Ensemble Methods	**15**	CO2, CO3, CO4	K3, K4

	Model Evaluation: Cross-Validation: K-fold Cross-Validation, Bias-Variance Tradeoff, Overfitting and Underfitting			
Module 3:	**Unsupervised Learning:** **Clustering:** K-means Clustering Algorithm, Hierarchical Clustering **Principal Component Analysis (PCA):** Introduction, Covariance Matrix and Eigenvalues, Dimensionality Reduction using PCA, Applications of PCA for High-Dimensional Data **Gaussian Mixture Models (GMM):** Introduction to GMM, Expectation-Maximization (EM) Algorithm, Model Selection in GMM	**15**	CO5	K3, K4
Module 4:	**Neural Networks and Reinforcement Learning:** **Neural Networks:** Introduction, Perceptron Model, Feedforward Neural Network Architecture, Activation Functions: Sigmoid, ReLU, Tanh, Gradient Descent and Backpropagation, Training Neural Networks: Cost Functions, Optimizers, **Deep Learning**: Introduction, Architecture of Deep Networks, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and LSTM **Reinforcement Learning,** Basic Concepts: Agent, Environment, Reward, Action, Q-learning and Temporal Difference Learning, Markov Decision Processes (MDP), Exploration vs. Exploitation	**15**	CO6	K2, K3, K4
Pedagogy:	Lectures/ Tutorials/Hands-on assignments//Flipped classroom			
Texts:	Alpaydin, E. (2020). Introduction to machine learning. MIT press, 4th Edition (or latest)			
References/ Readings:	1. Mitchell, T. M., (1997). Machine learning (Vol. 1, No. 9). New York: McGraw-hill. 2. Bishop, C. M. (2006). Pattern recognition and machine learning: springer New York.. 3. Flach, P. (2012). Machine learning: the art and science of algorithms that make sense of data. Cambridge university press. 4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. 5. Hart, Peter E., David G. Stork, and Richard O. Duda.(2000) Pattern classification. Hoboken: Wiley, 2000.			

	6. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
Web Resources:	1. MIT course on Introduction to Machine Learning. Retrieved on 27th May 2025, from: https://openlearninglibrary.mit.edu/courses/course-v1:MITx+6.036+1T2019/about

Title of the Course	Data Security, Privacy, and Ethics
Course Code	CSD-5006
Number of Credits	2
Theory/Practical	Theory
Level	500
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	CSD-5000	
Course Objectives:	To introduce foundational concepts of data privacy, ethics, and security, enabling students to responsibly manage and protect data in accordance with ethical principles and legal frameworks.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand fundamental concepts of data privacy, personal data, and ethical responsibility in data science.	PSO1, PSO2
	CO 2. Identify and evaluate ethical issues and privacy risks associated with data collection and usage.	PSO2, PSO3
	CO 3. Explain basic principles of data security and common types of security threats.	PSO1
	CO 4. Describe key data protection laws and compliance principles relevant to responsible data handling.	PSO7

Content:		No of hours	Mapped to CO	Cognitive Level
Module 1:	Introduction to Data Privacy: Definitions and significance, Personal vs. sensitive vs. anonymized data Personally Identifiable Information (PII): Identification and classification, Examples and consequences of misuse Data Lifecycle and Privacy Risks: Data collection, storage, processing, sharing, retention Ethical Foundations in Data Science: Ethical theories: Utilitarianism, Deontology, Virtue Ethics, Responsibilities of data scientists Fair Information Practices (FIPs): Transparency, accountability, consent, purpose limitation Privacy by Design and Default: Core principles and relevance in data projects Ethical Dilemmas and Real-world Case Studies: Cambridge Analytica, Aadhaar, social media surveillance Bias, Discrimination, and Fairness in Data Systems: Data-driven bias, algorithmic transparency, fairness metrics (introductory level)	**15**	CO1, CO2	K2, K3
Module 2:	Introduction to Data Security CIA Triad: Confidentiality, Integrity, Availability Risk, threat, and vulnerability concepts Common Security Threats: Malware, phishing, ransomware (conceptual) Real-world examples of data breaches Basic Security Mechanisms: Passwords, authentication, and access control, Introductory concept of encryption and hashing. Purpose of Data Protection Laws: Why laws are needed: protecting user rights and data responsibilities Key Legal Frameworks (Introductory)	**15**	CO3, CO4	K2, K3

	GDPR: Consent, data subject rights, accountability principles India's DPDP Act 2023: Applicability, personal data protection, user rights, HIPAA , Compliance and Ethical Responsibility, Basic role of a Data Protection Officer (DPO)			
Pedagogy:	Lectures/ Tutorials/Hands-on assignments//Flipped classroom			
Texts:	1. Davis, K. (2012). Ethics of big data: Balancing risk and innovation. O'Reilly Media. 2. Stallings, W. (2021). Information privacy engineering and privacy by design. Pearson. 3. Pfleeger, C. P., & Pfleeger, S. L. (2015). Security in computing (5th ed.). Pearson.			

Title of the Course	Data Modeling and Visualization
Course Code	CSD-5007
Number of Credits	2
Theory/Practical	Theory
Level	500
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	To analyze and interpret data throughout the data management lifecycle using appropriate analysis methods.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand and apply core data modeling concepts, including entities, relationships, constraints, and cardinalities.	PSO1, PSO3
	CO 2. Analyze and model real-world data complexities using strong/weak entities and relationship hierarchies.	PSO5
	CO 3. Explain the role of data visualization in analysis and decision-making, and distinguish between visualization types.	PSO1, PSO5
	CO 4. Create effective visualizations for various data types using appropriate techniques and	PSO5

	best practices.			
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	Data modelling- basic data modeling concepts and terminology, data modeling building blocks- hierarchies for the entities, data model Constraints for your attributes: specify cross-entity dependencies through strong and weak entities - summary of real-world entity and attributes complexities. real-world complexities to relationships why relationship cardinality and complexities matter - build real-world complexities into data model relationships-define the maximum cardinality of a relationship -define the minimum cardinality of relationship -use crow's foot notation for minimum and maximum cardinality -summary of cardinality and complex relationships. Introduction to Data Visualization: What is Data Visualization? Importance in Data Analysis and Decision Making. Components of Data Visualization: Data, Visual Encoding, Interaction, and Context. Types of Data Visualization: Static vs. Interactive Visualizations. Common types of charts: Bar Charts, Line Charts, Pie Charts, Histograms, Scatter Plots, Heatmaps. Visualization Best Practices: Effective use of color, labels, scales, and legends. Avoiding common pitfalls (misleading scales, cluttered visuals).	**15**	CO1, CO2, CO3	K1, K2, K3, K4
Module 2:	Visualizing Univariate and Bivariate Data: Visualizing Univariate Data: Histograms, Box Plots. Distribution of Data: Skewness, Kurtosis. Visualizing Bivariate Data: Scatter Plots, Heatmaps. Correlation, Trend Lines, and Regression. Introduction to Pair Plots and Correlation Matrices. Advanced Visualizations: Line Plots, Area Charts, and Time Series Decomposition. Understanding Trends, Seasonality, and Noise in Time Series Data. Geospatial Data Visualization: Mapping data with Geographic Information Systems (GIS). Visualizing location-based data using choropleth maps and bubble maps.	**15**	CO3, CO4	K1, K2, K3, K4

Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom
Texts:	1. Connolly, T., & Begg, C. (6th edition). Database systems: A practical approach to design, implementation, and management. Pearson Education. 2. Healy, K. (2018). Data visualization: A practical introduction. Princeton University Press.
References/ Readings:	1. Hoberman, S. (2009). Data modeling made simple: A practical guide for business and IT professionals (2nd ed.). Technics Publications. 2. Tufte, E. R. (2001). The visual display of quantitative information (2nd ed.). Graphics Press. 3. Tufte, E. R., Goeler, N. H., & Benson, R. (1990). Envisioning information (Vol. 2). Graphics Press. 4. Fry, B. (2008). Visualizing data: Exploring and explaining data with the processing environment. O'Reilly Media, Inc.

Title of the Course	Optimization Techniques
Course Code	CSD-5008
Number of Credits	4
Theory/Practical	Theory
Level	500
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	To provide students with a comprehensive understanding of Optimization techniques and their practical applications in decision-making.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Understand the fundamental concepts of Operations Research.	PSO1
	CO 2. Formulate real-world problems using linear programming models.	PSO3, PSO4
	CO 3. Apply various linear programming techniques in constrained environments.	PSO3
	CO 4. Solve transportation and assignment problems using standard methods.	PSO3
	CO 5. Analyze and solve problems using different game theory approaches.	PSO3
	CO 6. Apply operations research techniques to address practical industrial problems, aligning	PSO6

	theoretical models with industry perspectives.			
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	**Introduction to Operations Research** Mathematical models, Scope and applications, Phases of Operation Research study, Characteristics, Limitations of Operation Research. **Linear Programming** Introduction, Properties of Linear Programming, Basic assumptions, Mathematical formulation of Linear Programming, Limitations or constraints, Methods for the solution of LP Problem, Graphical analysis of LP, Graphical LP, Maximization problem, Graphical LP Minimization problem.	**15**	CO1, CO3	K1, K2
Module 2:	**Linear Programming Models** Simplex Method, Basics of Simplex Method, Formulating the Simplex Method, Simplex Method with two variables, Simplex Method with more than two variables, Big M Method. **Dual Linear Programming** Introduction, Primal and Dual problem, Dual problem properties, Solution techniques of Dual problem, Dual Simplex method, Relations between direct and dual problem, Economic interpretation of Duality.	**15**	CO2, CO3	K2, K3
Module 3:	**Transportation and Assignment Models** Introduction, Transportation problem, Balanced, Unbalanced, Methods of basic feasible solution Optimal solution, MODI method, Assignment problem, Hungarian Method. **Network Analysis** Basic concepts, Construction of Network, Rules and precautions CPM and PERT Networks Obtaining critical path, Probability and cost consideration, Advantages of	**15**	CO4	K2, K3, K4

	Network.			
Module 4:	**Theory of Games** Introduction, Terminology, Two Person Zero-Sum game, Solution of games with saddle points and without saddle points, 2X2 games, dominance principle, mX2 and 2Xn games, Graphical method. **Industry Perspective** Research and Analytical problems on various applications of the industrial issues.	**15**	CO5, CO6	K2, K3, K4
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom			
Texts:	Gupta, P. K., & Hira, D. S. (2022). Introduction to Operations Research. S. Chand Publishing.			
References/ Readings:	1. J K Sharma (2007), Operations Research Theory & Applications, 3e, Macmillan India Ltd. 2. Maurice Solient, Arthur Yaspen, Lawrence Fridman, OR methods and Problems (2003), New Age International Edition. 3. P. SankaraIyer, (2008), Operations Research, Tata McGraw-Hill. 4. Philips, D. T. (2007). Operations research: Principles and practice. John Wiley & Sons, Incorporated. 5. S.D. Sharma (2000). Operations Research. Nath& Co., Meerut.			
Web Resources:	1. MIT course on Optimization Methods. Retrieved on 27th May 2025, from: Optimization Methods | Sloan School of Management | MIT OpenCourseWare			

Title of the Course	Machine Learning Techniques Lab
Course Code	CSD-5009
Number of Credits	2
Theory/Practical	Practical
Level	500
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	CSD-5002, CSD-5003	
Course Objectives:	To equip students with practical skills to implement, evaluate, and apply machine learning algorithms using real-world data, fostering analytical thinking and hands-on experience through tools and mini-projects.	
Course Outcomes:		**Mapped to PSO**
	CO 1. To develop hands-on proficiency in implementing core machine learning algorithms and evaluating their performance using real-world datasets.	PSO 1, PSO 2, PSO 4
	CO 2. To equip students with the ability to preprocess data and apply appropriate supervised and unsupervised learning techniques.	PSO 2, PSO 3, PSO 4
	CO 3. To enable practical understanding of dimensionality reduction techniques and model evaluation methods through coding exercises and experiments.	PSO 2, PSO 3, PSO 4
	CO 4. To foster problem-solving and model-building skills by engaging in a mini-project using	PSO 2, PSO 3, PSO 4,

	machine learning tools and libraries.		PSO 6, PSO 8	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	1. Implement and evaluate a Decision Tree classifier using a real-world dataset. 2. Train a Random Forest classifier and compare accuracy with a Decision Tree on the same dataset. 3. Perform Simple Linear Regression on a dataset and evaluate using MSE and R^2. 4. Implement Multiple Linear Regression and interpret the results. 5. Train a Logistic Regression model on a binary classification dataset and evaluate with accuracy and confusion matrix. 6. Implement a Support Vector Machine (SVM) for classification using linear kernel and analyze results. 7. Apply K-fold cross-validation to any one classification model and report accuracy variance across folds.	**30**	CO1, CO2	K3, K4, K5
Module 2:	1. Apply K-means clustering to a dataset and visualize the resulting clusters. 2. Perform dimensionality reduction using Principal Component Analysis and visualize the transformed data. 3. Fit a Gaussian Mixture Model (GMM) and compare with K-means clustering on the same dataset. 4. Build and train a feedforward neural network for digit classification using a standard dataset. 5. Implement a Convolutional Neural Network (CNN) and evaluate its performance on an image dataset. 6. Mini Project: Design and implement a machine learning based mini project on a real-world dataset, involving model training, evaluation, and visualization.	**30**	CO2, CO3, CO4	K3, K4, K5
Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom/Mini-project			
Texts:	Alpaydin, E. (2020). Introduction to machine learning. MIT press, 4th Edition (or latest)			

References/ Readings:	1. Mitchell, T. M., (1997). Machine learning (Vol. 1, No. 9). New York: McGraw-hill. 2. Bishop, C. M. (2006). Pattern recognition and machine learning: springer New York. 3. Flach, P. (2012). Machine learning: the art and science of algorithms that make sense of data. Cambridge university press. 4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
Web Resources:	1. Machine Learning tutorial retrieved on 27th May 2025 from: https://www.w3schools.com/python/python_ml_getting_started.asp 2. Machine Learning with Python retrieved on 27th May 2025 from: https://www.freecodecamp.org/learn/machine-learning-with-python/

Title of the Course	Data Modeling and Visualization Lab
Course Code	CSD-5010
Number of Credits	2
Theory/Practical	Practical
Level	500
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil			
Course Objectives:	To design effective visualizations using best practices and tools like GapMinder and Tableau for clear, impactful data communication.			
Course Outcomes:			**Mapped to PSO**	
	CO 1. Design entity-relationship diagrams with proper entities, attributes, and cardinalities.		PSO1, PSO2	
	CO 2. Convert data models into normalized relational schemas using database tools..		PSO2, PSO3	
	CO 3. Create effective visualizations for univariate, bivariate, and time-series data.		PSO2, PSO5	
	CO 4. Interpret and communicate data insights using best visualization practices.		PSO8	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**

Module 1:	1. To crreate Entity-Relationship (ER) Diagrams using Crow's Foot notation; Practice defining strong and weak entities. 2. To define and apply cardinality (min/max) constraints in relationships; Model real-world complexities. 3. To translate ER diagrams into Relational Schemas; Normalize data to 3NF. 4. To use a database design tool (e.g., MySQL Workbench, dbdiagram.io) to create and visualize models.	**30**	CO1, CO2	K3, K4, K6
Module 2:	1. Introduction to visualization tools (e.g., Tableau, Power BI, or Python libraries like matplotlib/seaborn). 2. To create basic charts: bar charts, line charts, pie charts, histograms using sample datasets. 3. to visualize univariate data: Histograms, box plots; Understand skewness and kurtosis visually. 4. to visualize bivariate data: scatter plots, heatmaps; Explore correlation and trend lines.	**30**	CO3, CO4	K3, K4, K6
Pedagogy:	Tutorials/Hands-on Lab assignments/Flipped classroom/Mini-project			
Texts:	1. Connolly, T., & Begg, C. (6th edition). Database systems: A practical approach to design, implementation, and management. Pearson Education. 2. Healy, K. (2018). Data visualization: A practical introduction. Princeton University Press.			
References/ Readings:	1. Hoberman, Steve. Data modeling made simple: a practical guide for business and IT professionals. Technics Publications, 2nd Edition 2009. 2. Edward Tufte, The Visual Display of Quantitative Information 2nd Edition, 2001 3. Tufte, Edward R., Nora Hillman Goeler, and Richard Benson. Envisioning information. Vol. 2. Cheshire, CT: Graphics press, 1990. 4. Fry, Ben. Visualizing data: Exploring and explaining data with the processing environment. " O'Reilly Media, Inc.", 1st Edition 2008.			

Discipline Specific Elective Courses

Title of the Course	Domain-Specific Predictive Analytics
Course Code	CSD-5203
Number of Credits	4
Theory/Practical	Theory
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	To provide a comprehensive theoretical foundation in predictive modeling techniques, with a focus on their application across various real-world domains.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Develop a strong foundation of the core concepts of predictive modeling.	PSO1
	CO 2. Understand the key features for predictive models for financial data.	PSO3, PSO4
	CO 3. Compare the machine learning models for clinical prediction and fraud detection in Healthcare.	PSO3, PSO4
	CO 4. Use the predictive models for customer and market analysis.	PSO3, PSO4

	CO 5. Gain insights on the application of machine learning models in social media analytics.		PSO3, PSO4	
Content:		**No of hours**	**Mapped to CO**	**Cognitive Level**
Module 1:	**Introduction** Prediction vs Interpretation, Key ingredients of predictive models, Predictive modeling process, Linear and Non-linear regression models, Linear and Non-linear classification models, Class imbalance problem. **Finance Domain** Financial Data Structures, Labeling and Meta-Labeling, Sample Weights, Structural Breaks, Entropy Features, Microstructural Features, Ensemble Methods, Cross-Validation in Finance, Feature Importance, Hyper-Parameter Tuning, Backtesting.	**15**	CO1, CO2	K2, K3
Module 2:	**Healthcare Domain** Healthcare Data Sources, Challenges in Healthcare Data Analysis, Clinical Prediction Models - Linear Regression, Logistic Regression, Bayesian Models, Neural Networks, Cost-Sensitive Learning, Multiple Instance Learning, Sparse Methods, Kernel Methods, Fraud detection in Healthcare.	**15**	CO3	K2, K3
Module 3:	**Marketing Domain** Understanding Markets, Building customer profiles, Managing customer portfolio, Predicting customer personas, Predicting customer journey, Predicting customer value, Predicting likelihood to buy, Predicting customer choice, Targeting current customers, Finding new customers, Retaining customers, Positioning and Promoting products, Recommending products.	**15**	CO4	K2, K3
Module 4:	**Social Media Domain** Social media prediction, Network measures, Network models, Properties of Real-World Networks, Modeling Real World Networks with Random Graphs, Small World Model, Preferential Attachment Model, Community Detection, Evolution and Evaluation, Information diffusion in Social Media.	**15**	CO5	K2, K3

Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom
Texts:	1. Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26, p. 13). New York: Springer. 2. De Prado, M. L. (2018). Advances in financial machine learning. John Wiley & Sons. 3. Reddy, C. K., & Aggarwal, C. C. (Eds.). (2015). Healthcare data analytics. CRC Press. 4. Artun, O., & Levin, D. (2015). Predictive marketing: Easy ways every marketer can use customer analytics and big data. John Wiley & Sons. 5. Zafarani, R., Abbasi, M. A., & Liu, H. (2014). Social media mining: an introduction. Cambridge University Press.
References/ Readings:	1. Klaas, J. (2019). Machine learning for finance: principles and practice for financial insiders. Packt Publishing Ltd. 2. Miller, T. W. (2015). Marketing data science: Modeling techniques in predictive analytics with R and Python. FT Press. 3. Shu, K., & Liu, H. (2022). Detecting fake news on social media. Springer Nature. 4. Hong, T. P., Serrano-Estrada, L., Saxena, A., & Biswas, A. (Eds.). (2022). Deep Learning for Social Media Data Analytics. Springer International Publishing.
Web Resources:	MIT course on Machine Learning for Healthcare. Retrieved on 28th May 2025, from: Machine Learning for Healthcare | Electrical Engineering and Computer Science | MIT OpenCourseWare

Title of the Course	Web Data Analytics
Course Code	CSD-5204
Number of Credits	4
Theory/Practical	Theory
Level	400
Effective from AY	2025-26
New Course	Yes
Bridge Course/ Value added Course	No
Course for advanced learners	No

Pre-requisites for the Course:	Nil	
Course Objectives:	The course will help the learner to make strategic decisions based on customer interactions and business intelligence on the web.	
Course Outcomes:		**Mapped to PSO**
	CO 1. Learner will understand in basic concept of web analysis & analytics, while also understanding the relevant web technologies	PSO1, PSO2
	CO 2. Learner will understand and apply the various methods & sources for web data collections	PSO2, PSO6
	CO 3. Learner will apply various methods qualitative analysis and quantitative measures	PSO1, PSO3, PSO4
	CO 4. Learner will understand the various analytics aspects that will generate insights from web data collected, for the purpose of strategic decision making	PSO3, PSO5, PSO8

Content:		No of hours	Mapped to CO	Cognitive Level
Module 1:	Introduction & Relevant Technologies ● Definition, Process, Key terms & Key phrases ● Building blocks of web analytics ● Offsite web, On site web; Web analytics platform ● Internet & TCP/IP, Client / Server Computing, HTTP, Server Log Files & Cookies, Web Bugs	**15**	CO1	K1, K2
Module 2:	Data Collection & Qualitative Analysis ● Data Collection via Clickstream Data, Outcomes Data, Research data & Competitive Data ● Heuristic evaluations; site visits ● Website & post-visit Surveys	**15**	CO2, CO3	K2, K3
Module 3:	Analytic Fundamentals & Using Web Metrics ● Capturing data via web logs, javascript tags, etc. ● Separating data serve & data capture ● Link coding issues ● Common page metrics (page view, hits, unique visitors, average time on website) ● Gauging optimization metrics (bounce rate, conversion rate, etc.) ● Reports (real-time, average traffic, etc.) ● KPI; perspectives	**15**	CO2, CO3, CO4	K2, K3, K4
Module 4:	Web Analytics 2.0 & Google Analytics ● Overview of Web Analytics of 1.0 & 2.0 ● Competitive intelligence analysis ● Website traffic analysis ● Google Analytics; Adwords; benchmarking ● Google website optimizer; Paid & Organic traffic; privacy concerns	**15**	CO3, CO4	K2, K3, K4, K5

Pedagogy:	Lectures/ Tutorials/Hands-on assignments/Flipped classroom
Texts:	1. Clifton, B. (2012). Advanced web metrics with Google Analytics. John Wiley & Sons. 2. Kaushik, A. (2009). Web analytics 2.0: The art of online accountability and science of customer centricity. John Wiley & Sons.
References/ Readings:	Sterne, J. (2003). Web metrics: Proven methods for measuring web site success. John Wiley & Sons.
Web Resources:	1. Open Web Analytics. (n.d.). Open Web Analytics. Retrieved May 9, 2025, from https://www.openwebanalytics.com/ 2. Google. (n.d.). Google Analytics Help. Google. Retrieved May 9, 2025, from https://support.google.com/analytics