CENTRAL UNIVERSITY OF HARYANA

(Established under the Central Universities Act, 2009)

(NAAC Accredited 'A' Grade)

NEP and NCrF/NHEQF based

Curriculum and Syllabi of

M. Sc. (Data Science)

(w.e.f. 2025-26)

Department of Computer Science and Information Technology

School of Basic Sciences

Table of Contents

1. Vision and Mission

i) **Vision and Mission of the University**

Vision

To develop enlightened citizenship of a knowledge society for peace and prosperity of individuals, nations, and the world, through the promotion of innovation, creative endeavors, and scholarly inquiry.

Mission

To serve as a beacon of change, through multi-disciplinary learning, for the creation of a knowledge community, by building a strong character and nurturing value-based transparent work ethics, promoting creative and critical thinking for holistic development and self- sustenance for the people of India. The University seeks to achieve this objective by cultivating an environment of excellence in teaching, research, and innovation in pure and applied areas of learning.

ii) Vision and Mission of the Department
Vision

To be a Centre of Excellence for nurturing computer professionals with strong application expertise through experiential learning and research for matching the requirements of industry and society instilling in them the spirit of innovation and entrepreneurship by providing knowledge of computer systems in both hardware and software application design so that they contribute not only in the progress of software and its application but even encompass the entire emerging domain of computer technology.

Mission

1. To improve high-quality professional training at the postgraduate with an Emphasis on the basic principle of Computer Science and application. To impart value-based, quality education that provides design and development like software applications in their entirety. Innovative learning-centric facilities for solving computational problems.
2. To promote research-based activities through analysis and interpretation of data and synthesis of the information for utilization in resolving practical problems relating to computer applications.

3. To provide help in promote\preparing students to qualify for exams like UGC-NET, GATE, and other competitive exams.
4. To provide a framework through Project Based Learning to support society and industry in promoting a multidisciplinary activity.
5. To provide a quality learning experience through effective classroom practices, the active learning style of teaching, and opportunities for meaningful interaction between students and faculty.
6. To develop a crystal-clear evaluation system and experiential learning mechanism aligned with futuristic technologies and industry.
7. To undertake societal activities for the upliftment of rural/deprived sections of the society.

iii) Mapping of Vision and Mission

Vision and Mission of the University	**Vision and Mission of the Department**
To develop enlightened citizenship of a knowledge society for peace and prosperity of individuals, nations, and the world, through the promotion of innovation, creative endeavors, and scholarly inquiry.	**Yes**
To serve as a beacon of change, through multi-disciplinary learning, for the creation of a knowledge community, by building a strong character and nurturing value-based transparent work ethics, promoting creative and critical thinking for holistic development and self-sustenance for the people of India.	**Yes**
The University seeks to achieve this objective by cultivating an environment of excellence in teaching, research, and innovation in pure and applied areas of learning.	**Yes**

2. Background

i) NEP-2020 and LOCF an integrated Approach

In alignment with the National Education Policy (NEP) 2020 and the UGC's Learning Outcome-Based Curriculum Framework (LOCF), the Central University of Haryana undertook a comprehensive curriculum revision process for its undergraduate and postgraduate programmes, including the M. Sc.. Data Science programme. This initiative was driven by the adoption of the Comprehensive Roadmap for Implementation of NEP-2020, approved during the 32nd Academic Council Meeting held on April 23, 2021.

The curriculum of M. Sc.. Data Science has been restructured to reflect the interdisciplinary, holistic, and outcome-oriented approach promoted by NEP-2020. Faculty members were oriented through workshops and webinars to facilitate the integration of core NEP features, including:

- Multidisciplinary learning and flexibility in course choices.
- Integration of **21st-century skills**, including **data-driven decision-making**, **computational thinking**, **problem-solving**, and **ethical AI**.
- Focus on **student-centric** and **experiential learning**, such as **project-based**, **discovery-based**, and **discussion-based** pedagogy.
- Emphasis on **blended learning** with a 60:40 ratio (face-to-face and online), enabling access to the latest tools and techniques in data science.
- Promoting exposure to **Indian knowledge systems**, values-based learning, and **Global Citizenship Education (GCED)**.
- Industry collaborations and internship opportunities to bridge the gap between academic learning and real-world applications in data science, analytics, and AI.

The revised curriculum was formulated through an inclusive process involving faculty, departments, and external experts. A uniform curriculum structure was adopted that includes: background, programme outcomes, programme-specific outcomes, graduate attributes, course structure, learning outcomes, pedagogical methods, assessment, and evaluation frameworks.

ii) About subject

Data Science has emerged as a transformative discipline that blends **statistics, computer science, and domain-specific knowledge** to extract actionable insights from data. With exponential growth in digital data, data science plays a pivotal role in sectors like **healthcare, finance, marketing, education, public policy, environment**, and more.

M. Sc. Data Science integrates modern computational tools with robust statistical techniques to prepare students to handle real-world data problems. The subject covers key areas such as:

- Data Wrangling and Visualization
- Machine Learning and AI
- Big Data Technologies
- Cloud Computing
- Statistical Inference and Predictive Modeling
- Deep Learning and Natural Language Processing

The programme emphasizes **hands-on training** through tools like **Python, R, SQL, Hadoop, Spark, TensorFlow**, and cloud platforms (e.g., AWS, Azure). Students are also trained in **ethical data handling**, **data governance**, and **privacy-preserving methods**, which are essential in today's data-centric world.

- Enhances creativity & thinking skills
- Proves beneficial for career aspiration
- Design and develop a software application for different industries
- Provides efficient & better use of IT Technology
- Improves research work & helps in communicating with different education providers
- Gives instant information/ Quick processing of data on any topic in just a single click
- Manages the software, hardware & networks in any industry
- Involves in the design and development of the hardware components of PCs & laptops
- Develop software for peripheral computing devices such as printers, modems, scanners, etc.
- Write code and algorithms for operating systems like Windows, Linux, etc.
- Develop design, implementation, and management of information systems of computer

hardware and software.

iii) About the Programme (Nature, extent, and aims):

The M. Sc. Data Science (2-year) **programme is designed to offer a** balanced blend of theoretical foundations and practical applications **in data analytics, machine learning, and computational techniques.**

Objectives:

- Equip students with strong analytical and programming skills for solving data-centric problems.
- Foster interdisciplinary thinking by combining data science with domain expertise.
- Promote research, innovation, and entrepreneurship in data-driven domains.
- Bridge the academia-industry gap through industry-driven projects, internships, and mentorship.

The programme structure includes **core courses**, **electives**, **practical labs**, and a **major research/project component** in the final semester. Continuous assessment, capstone projects, and industry participation ensure relevance and rigour.

iv) Qualification Descriptors (possible career pathways)

Graduates of M. Sc. Data Science will be well-equipped for diverse roles in both industry and academia. Possible career pathways include:

1. **Data Scientist**

   Analyze and interpret complex data to assist in strategic decision-making.

2. **Machine Learning Engineer**

   Design and develop ML models for automation and predictive analysis.

3. **Data Analyst / Business Intelligence Analyst**

   Create dashboards and reports; support data-driven decision-making.

4. **AI Specialist**

   Develop algorithms for AI-powered applications in various domains.

5. **Big Data Engineer**

   Work on scalable data infrastructure using tools like Hadoop, Spark, Kafka.

6. **Research Scientist**

   Contribute to academic or industrial R&D in emerging areas of data science.

7. **Data Engineer**

   Manage data pipelines and databases for efficient data storage and retrieval.

8. **Statistical Analyst**

   Apply statistical methods to extract insights from experimental or survey data.

9. **Cloud Data Specialist**

   Develop and manage data systems on cloud platforms (e.g., AWS, GCP, Azure).

10. **Higher Education and Research**

    Pursue Ph.D. or M. Tech. in AI, Data Science, Computational Sciences, or related fields.

The programme also promotes **entrepreneurship** and **innovation**, encouraging graduates to develop data-driven products or services that address societal and business challenges.

3. Program Educational Objectives (PEOs)

PEO 1: Core Competency Development

- To develop strong foundational knowledge in statistics, mathematics, computer science, and data science tools to solve complex real-world problems through data-driven approaches.

PEO 2: Professional Excellence and Lifelong Learning

- To excel in professional careers or higher education by continuously updating their knowledge and skills in emerging technologies such as machine learning, AI, big data analytics, and cloud computing.

PEO 3: Innovation, Research, and Entrepreneurship

- To demonstrate innovation and research capabilities by contributing to the development of new analytical methods, data products, or startups that address societal and industrial challenges.

PEO 4: Ethical, Social, and Environmental Responsibility

- To uphold ethical standards and contribute to society by applying data science responsibly, considering privacy, fairness, sustainability, and the broader impact of their solutions.

PEO 5: Communication, Collaboration, and Leadership

- To effectively collaborate in multidisciplinary teams, communicate technical information clearly, and take on leadership roles in academia, industry, or public sector organizations.

4. Programme Outcomes (POs)

Students enrolled in the Master Programmes offered by the Departments under the School of Basic Sciences will have the opportunity to learn and master the following components in addition to attaining important essential skills and abilities:

Sr. No.	Component	Outcomes
PO-1	Computational Knowledge:	Apply knowledge of computing fundamentals, computing specialization, mathematics, and domain knowledge appropriate for the computing specialization to the abstraction and conceptualization of computing models from defined problems and requirements.
PO-2	Problem Analysis:	Identify, formulate, research literature, and solve complex computing problems reaching substantiated conclusions using fundamental principles of mathematics, computing sciences, and relevant domain disciplines.
PO-3	Design /Development of Solutions	Design and evaluate solutions for *complex* computing problems, and design and evaluate systems, components, or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.
PO-4	Conduct Investigations of Complex Computing Problems	Use research-based knowledge and research methods including design of experiments, analysis, and interpretation of data, and synthesis of the information to provide valid conclusions.
PO-5	Modern Tool Usage:	Create, select, adapt and apply appropriate techniques, resources, and modern computing tools to *complex* computing activities, with an understanding of the limitations.

PO-6	Professional Ethics	Understand and commit to professional ethics and cyber regulations, responsibilities, and norms of professional computing practice.
PO-7	Life-long Learning	Recognize the need, and have the ability, to engage in independent learning for continual development as a computing professional.
PO-8	Project management and finance	Demonstrate knowledge and understanding of the computing and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
PO-9	Communication Efficacy	Communicate effectively with the computing community, and with society at large, about *complex* computing activities by being able to comprehend and write effective reports, design documentation, make effective presentations, and give and understand clear instructions.
PO-10	Societal and Environmental Concerns	Understand and assess societal, environmental, health, safety, legal, and cultural issues within local and global contexts, and the consequential responsibilities relevant to professional computing practice.
PO-11	Individual and Team Work	Function effectively as an individual and as a member or leader in diverse teams and in multidisciplinary environments.
PO-12	Innovation and Entrepreneurship	Identify a timely opportunity and use innovation to pursue that opportunity to create value and wealth for the betterment of the individual and society at large.

5. Programme Specific Outcomes (PSOs)

The postgraduates shall be able to realize the following outcomes by the end of program studies:

Number	Programme Specific Outcomes
PSO-1	The ability to remember and understand the basic concept of associated subjects and Computer Fundamentals, Computer Programming, Design, and Analyze different Network Techniques.
PSO-2	The proficiency to understand, evaluate and analyse the design and algorithm concepts of computer architecture, Operating systems, Computer Networks, Software Engineering, Design and Analysis of Algorithms, Compiler Design, Artificial Intelligence, etc
PSO-3	The ability to design and solve problems in the field of Interdisciplinary subjects by applying the knowledge acquired from Data analysis, Software development & other allied topics.
PSO-4	The skills to develop, adopt, and assess the latest innovative industry best practices, then analyze and comprehend the young mindsets accordingly to their attitude toward higher studies, research, and to possess a successful path as a young entrepreneur.
PSO- 5	Analyze their abilities in systematic planning, developing, testing, and executing complex computing applications, in the field of Social Media and Analytics, Web Application Development, and Data Interpretations.

6. Postgraduate Attributes

On completion of the post-graduate programme in M Sc, students are expected to equip with the skills of creative, critical, and rational thinking associated with computers and their use for human society. The following attributes are expected from the students of M Sc:

Sr. No.	P.G. Attributes
PGA-1	Describe the notion of mathematical thinking, mathematical proofs, and algorithmic thinking, and be able to apply them in problem-solving that involves Discrete Mathematical Structures, Design and Analysis of Algorithms, image processing, Compiler Design, etc.
PGA-2	Ability to use the updated tools, techniques, and modern Software tools necessary for software Development like Android Application Development, Data Science with R programming, Bioinformatics, Cloud Computing, etc.
PGA-3	Introduce the basic principles, techniques, and applications of Artificial Intelligence. Emphasis will be placed on the teaching of these fundamentals and labs for the _hands-on‘ approach to understanding, as well as a challenging avenue for exploration and creativity.
PGA-4	Provide professional knowledge in specialized areas such as Computer Vision, Internet of Things, Natural Language Processing, Speech Recognition, etc.
PGA-5	Communicate effectively by comprehending, documenting, making effective presentations, and exchanging clear instructions through project reports and presentations.
PGA-6	Describe the fundamental concepts, Solve problems, use algorithms in machine learning and popular machine learning algorithms with programming in Python/MATLAB. And describe the concept of Deep Learning.

PGA-7	Design and implement smart, intelligent, and user-friendly interfaces for computer web applications using PHP version 5. Students will learn how to connect to any ODBC-compliant database and perform hands-on practice with a MySQL database to create database-driven HTML forms and reports, etc. Students also learn how to configure PHP and Apache Web Server.
PGA-8	Develop programming skills through C, PHP, object-oriented programming, Java Programming, Advance Java Programming & Android, etc.
PGA-9	Develop robust, efficient software systems, Analyse the security requirements, examine software problems in multiple aspects including, testing and implementation that would involve software engineering, data structures, networked programming, Wireless Sensor Networks, Network Security, etc
PGA-10	Analyze the relationship between organizations, information systems, business processes, Identify different techniques, store, manipulate the huge data from different resources, involving E-commerce and Data Science with R programming.

7. Mapping to SDGs

The **Sustainable Development Goals (SDGs)**, adopted by the United Nations in 2015 as part of the **2030 Agenda for Sustainable Development**, consist of **17 global goals** designed to end poverty, protect the planet, and ensure peace and prosperity for all. Education plays a central role across many of these goals, and is specifically highlighted in **SDG 4**.

Key Targets of SDG 4 include:

- Universal primary and secondary education
- Early childhood development and pre-primary education
- Equal access to affordable technical, vocational, and higher education
- Elimination of gender disparities
- Promotion of education for sustainable development and global citizenship
- Improving literacy and numeracy
- Building and upgrading education facilities that are child-, disability-, and gender-sensitive

List of SDGs

SDG1: No Poverty- *End poverty in all its forms everywhere.*

SDG2: Zero Hunger- *End hunger, achieve food security and improved nutrition, and promote sustainable agriculture.*

SDG3: Good Health and Well-being- *Ensure healthy lives and promote well-being for all at all ages.*

SDG4: Quality Education- *Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.*

SDG5: Gender Equality- *Achieve gender equality and empower all women and girls.*

SDG6: Clean Water and Sanitation- *Ensure availability and sustainable management of water and sanitation for all.*

SDG7: Affordable and Clean Energy- *Ensure access to affordable, reliable, sustainable and modern energy for all.*

SDG8: Decent Work and Economic Growth- *Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.*

SDG9: Industry, Innovation and Infrastructure- *Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation.*

SDG10: Reduced Inequality- *Reduce inequality within and among countries.*

SDG11: Sustainable Cities and Communities -*Make cities and human settlements inclusive, safe, resilient, and sustainable.*

SDG12: Responsible Consumption and Production -*Ensure sustainable consumption and production patterns.*

SDG13: Climate Action- *Take urgent action to combat climate change and its impacts.*

SDG14: Life Below Water- *Conserve and sustainably use the oceans, seas, and marine resources.*

SDG15: Life on Land- *Protect, restore and promote sustainable use of terrestrial ecosystems, manage forests sustainably, combat desertification, and halt biodiversity loss.*

SDG16: Peace, Justice and Strong Institutions - *Promote peaceful and inclusive societies for sustainable development, provide access to justice for all, and build effective, accountable institutions.*

SDG17: Partnerships for the Goals - *Strengthen the means of implementation and revitalize the global partnership for sustainable development.*

Table: Mapping of SDG to Courses

Sr. No.	Course Title	SDGs Mapped	Rationale
1.	Artificial Intelligence	4, 8, 9	Equips learners with cutting-edge skills in machine learning, data science, robotics, and automation
2.	Data Structures and Algorithms	4	Enhances Computational Thinking, Builds Employable Technical Skills
3.	Mathematics for Data Science	4, 13, 16	Develops Analytical Thinking, Supports Career & Academic Growth
4.	Artificial Intelligence Lab using Python	4, 8, 9	Promotes problem-solving and innovation, Enhance scientific research, upgrade the technological capabilities of industrial sectors, Promotes entrepreneurship through AI-driven product development
5.	Data Structures and Algorithms Lab	4, 8	Skill development, Boosts productivity

6.	Information Systems Security	4, 5, 8, 9	Equips learners with in-demand cybersecurity skills, Drives secure digital growth and high-value employment, Builds resilient and secure digital infrastructures, Strengthens digital governance, public trust, and national security.
7.	Optimization Techniques	4, 8, 9, 12, 13	Drives efficiency and innovation in systems and industries; Equips learners with quantitative problem-solving skills; Supports sustainable production and climate-resilient planning; Contributes to economic and industrial modernization
8.	Data Warehousing and Data Mining	3, 4, 8, 9	Enhances employability in data-intensive industries, Builds core skills in managing and analyzing large-scale data, Supports research, healthcare analytics, and smart city planning, Powers innovation through data-driven insights
9.	Cloud Computing	13, 17	Supports sustainable computing through efficient infrastructure, Facilitates global cooperation and knowledge sharing through cloud-hosted platforms
10.	Database Management Systems	4, 8, 9, 16	Prepares students for careers in data-driven roles such as data analyst, database administrator, and backend developer.
11.	Machine Learning Techniques	3, 4, 8, 9	Enables smart health solutions; Builds cutting-edge skills in data science and AI; Supports innovation across sectors; Increases career readiness and economic opportunities
12.	Database Management Systems Lab	4, 5, 8, 9	Offers hands-on, technical skill development, Prepares students for roles in data-driven industries, Supports digital infrastructure and innovation through database expertise
13.	Machine Learning Techniques Lab	4, 9, 11, 13, 16	4, 9, 11, 13, 16
14.	Natural Language Processing	3, 4, 8, 9, 16, 17	Enhances access to information across languages and regions; Promotes inclusive, equitable, and multilingual digital systems; Supports automation and innovation in industry and healthcare; Builds skills and opportunities in AI-driven global job markets
15.	Data Science for Internet of Things	4, 9, 11, 12, 13	Powers real-time, intelligent systems for smarter infrastructure;

			Enables efficient resource usage and sustainable living environments; Trains learners in next-generation technologies; Supports climate action and eco-friendly innovation
16.	Time Series Analysis and Forecasting	4, 8, 9, 13	Powers predictive systems in health, climate, economy, and industry; Strengthens resilience and planning through data trends and forecasting; Builds modern analytical and technical competencies in students; Drives innovation and sustainable practices in multiple domains
17.	Big Data Analytics	3, 9, 12, 13	Empowers data-driven innovation and sustainability; Enables predictive, prescriptive, and descriptive analytics for better decisions; Fosters economic growth and high-skill employment; Supports climate, health, and education solutions through data
18.	Deep Learning Techniques	3, 9, 12, 13	Drives AI-powered innovation across industries; Enables breakthroughs in healthcare, environment, and smart cities; Trains students in advanced computational and cognitive technologies; Enhances economic productivity and supports sustainability efforts
19.	Data Visualization	3,4, 9, 16	Makes data accessible, interpretable, and actionable; Empowers stakeholders in education, governance, and industry; Enhances transparency, communication, and informed decision-making; Bridges the gap between raw data and real-world impact
20.	Big Data Analytics Lab	3, 9, 12, 13	Empowers data-driven innovation and sustainability; Enables predictive, prescriptive, and descriptive analytics for better decisions; Fosters economic growth and high-skill employment; Supports climate, health, and education solutions through data
21.	Deep Learning Techniques Lab	3, 9, 12, 13	Drives AI-powered innovation across industries; Enables breakthroughs in healthcare, environment, and smart cities; Trains students in advanced computational and cognitive technologies; Enhances economic

			productivity and supports sustainability efforts
22.	Dissertation Seminar	4, 8, 9, 17	Applies technical knowledge in meaningful and practical ways, Builds innovation, problem-solving, and employability, Encourages collaborative, research-driven solutions aligned with real-world needs, Potential to address or support specific SDGs depending on project themes.
23.	Summer Training/ Internship	4, 8, 9, 17	Enhances job-readiness and real-world skills; Promotes technical innovation and workplace productivity, Builds bridges between academia and industry, Empowers youth to contribute to a sustainable economy
24.	Blockchain Technology and Applications	4, 9, 16, 17	Enables trust, transparency, and accountability in digital systems; Powers cross-sector innovation through decentralization; Enhances career readiness and modern tech skills; Supports secure and inclusive digital economies
25.	Quantum Computing	3, 13, 17	Supports healthcare breakthroughs through molecular modeling; Holds potential to tackle climate and sustainability challenges, Encourages international collaboration in cutting-edge science
26.	Software Engineering	4, 8, 9, 17	Develops critical skills for the digital workforce, Enables innovation and modern digital services, Drives entrepreneurship and job creation, Promotes global collaboration and tech-enabled SDG solutions
27.	Dissertation	4, 8, 9, 17	Encourages hands-on learning and innovation; Prepares students for industry and research careers; Builds teamwork, creativity, and responsibility; Can directly address real-world SDG challenges

8. Structure of Master of Science (Data Science)

General, Course structure & Theme & Semester-wise credit distribution

A. Definition of Credit:

1 Hour Lecture (L) per week	1 credit
1 Hour Tutorial (T) per week	1 credit
2 Hours Practical (Lab)/week	1 credit

B. Total credits:

The total credits of the M Sc. Data Science programme are 88. The minimum qualifying marks for each course of the programme shall be 40%.

C. Structure of M. Sc. program:

Sr. No.	Category	Breakup of Credits	%
1	**Professional Core Courses**	46	52.27
2	**Program Elective Courses**	12	13.64
3	**General Elective Courses**: (Taken from other departments)	8	9.09
4	**Project work** and Internship in Industry / in house	1+1+20=22	25.0
5	**Mandatory Courses as bridge course:** [Computer Fundamentals and C Programming]	Non- Credit	
	Total	88	

9. Learning Outcome Index

(Mapping of Courses with POs and PSOs)

Mapping of Courses with POs (First Year)

Semester	POs ⇨ / Courses ⇩	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6	PO-7	PO-8	PO-9	PO-10	PO-11	PO-12
I	CC-1	✓	✓	×	✓	×	×	✓	×	✓	✓	×	×
	CC-2	✓	✓	×	×	✓	×	✓	×	✓	×	✓	×
	CC-3	✓	✓	×	✓	×	×	✓	×	✓	×	×	×
	CCP-1	✓	✓	×	✓	✓	×	✓	✓	✓	✓	×	✓
	CCP-2	✓	✓	✓	✓	×	✓	✓	×	✓	×	✓	✓
	DSE - 1	✓	×	×	×	✓	×	✓	✓	✓	×	✓	×
	DSE - 2	✓	×	×	×	✓	×	✓	✓	✓	×	✓	×
	DSE - 3	✓	×	×	×	✓	×	✓	✓	✓	×	✓	×
	BC- 1	✓	✓	×	✓	✓	×	✓	×	✓	✓	✓	✓
	MDC-1	✓	✓	×	✓		×	✓	×	✓	✓	✓	✓
II	CC-1	✓	✓	×	×	✓	×	✓	✓	✓	✓	✓	✓
	CC-2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
	CC-3	✓	✓	×		✓	×	✓	✓	✓	✓	✓	✓
	CCP-1	✓	✓	×	✓	✓	×	✓	×	✓	✓	×	✓
	CCP-2	✓	✓	×	✓	✓	✓	✓	×	✓	×	✓	✓
	DSE -1	✓	×	×	×	✓	×	×	×	✓	✓	×	×
	DSE -2	✓	×	×	×	✓	×	✓	×	✓	×	×	×
	DSE -3	✓	✓	×	×	✓	×	✓	×	✓	✓	×	×
	MDC-2	✓	×	✓	×	✓	×	×	×	✓	✓	✓	✓

Mapping of Courses with POs (Second Year)

Semester	POs ⇨ / Courses ⇩	PO-1	PO-2	PO-3	PO-4	PO-5	PO-6	PO-7	PO-8	PO-9	PO-10	PO-11	PO-12
III	CC-1	✓	×	×	✓	✓	✓	✓	×	✓	✓	✓	✓
	CC-2	✓	✓	×	✓	✓	×	×	×	✓	×	✓	×
	CC-3	✓	✓	×	✓	✓	×	✓	×	✓	×	×	✓
	CCP-1	✓	✓	×	✓	✓	✓	×	✓	✓	✓	✓	✓
	CCP-2	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓
	DSE -1	✓	✓	×	✓	✓	✓	×	×	✓	✓	×	✓
	DSE -2	✓	×	×	✓	×	✓	×	×	✓	✓	×	✓
	DSE -3	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
	MP	✓	✓	✓	×	×	✓	✓	✓	✓	✓	✓	✓
	STR	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓
IV	CC-1	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓

Mapping of Courses with PSOs (First Year)

Semester	POs ⇨	PSO-1	PSO-2	PSO-3	PSO-4	PSO5
	Courses ⇩					
I	CC-1	✓	✓	✓	✓	✓
	CC-2	✓	✓	✓	✓	✓
	CC-3	✓	✓	✓	✓	✓
	CCP-1	×	✓	✓	✓	✓
	CCP-2	✓	✓	✓	✓	✓
	DSE- 1	✓	✓	✓	✓	✓
	DSE- 2	×	✓	✓	✓	✓
	DSE- 3	✓	×	×	✓	✓
	BC- 1	✓	✓	✓	✓	✓
	MDC-1	✓	✓	✓	✓	✓
II	CC-1	✓	✓	✓	✓	✓
	CC-2	✓	✓	✓	✓	✓
	CC-3	✓	✓	✓	✓	✓
	CCP-1	✓	✓	✓	✓	✓
	CCP-2	✓	✓	×	✓	✓
	DSE-1	✓	✓	✓	✓	✓
	DSE-2	×	✓	✓	✓	✓
	DSE-3	✓	✓	✓	✓	✓
	MDC-2	✓	✓	✓	✓	✓

Mapping of Courses with PSOs (Second Year)

Semester	POs ⇨	PSO-1	PSO-2	PSO-3	PSO-4	PSO-5
	Courses ⇩					
III	CC-1	✓	✓	✓	×	✓
	CC-2	✓	✓	✓	✓	✓
	CC-3	✓	✓	✓	×	✓
	CCP-1	✓	×	×	✓	✓
	CCP-2	✓	✓	✓	✓	✓
	DSE-1	✓	✓	✓	✓	✓
	DSE-2	✓	×	✓	✓	✓
	DSE-3	✓	✓	✓	✓	✓
	MP	✓	✓	✓	✓	✓
	STR	✓	✓	✓	✓	✓
IV	CC-1	✓	✓	✓	✓	✓

10. Semester-wise Courses and Credit Distribution

Total Credit = 88

Semester Wise Distribution of Credits: 22 + 22 + 24 +20 = 88

Eligibility for Admission:

Passed B.C.A/ B.Sc. (Computer Science)/ B.Sc. (IT) / B.E. (CSE)/ B.Tech. (CSE) / B.E. (IT) / B.Tech. (IT) or equivalent Degree.

OR

Passed any graduation degree (e.g.: B.E. / B.Tech. / B. Sc. / B.Com. / B.A./ B. Voc./ etc.,) with Mathematics at 10+2 level or at Graduation Level Obtained at least 50% marks (45% marks in case of candidates belonging to reserved category) in the qualifying examination. (The students admitted with this eligibility will have to simultaneously undertake additional Bridge Course(s) as prescribed by the University during the first year).

* *Reservation policy, Fee etc. as per CU Haryana rules.*

Intake: 30

Category-wise Credit distribution

Category	**Credits**
• Programme core	68
• Departmental elective	12
• General elective	08

Bridge Course (0-credit)

Sr. No.	**Course Code**	**Course Name**	**L**	**T**	**P**	**Credits**
1.	CST401BC00	Computer Fundamentals and C Programming	**3**	**0**	**2**	0

Note: It is mandatory for all students from **non-computer backgrounds** to pass out Bridge Course within the first year. Bridge Course will be offered only in the first semester and is common to students enrolled in MCA and M.SC. (Data Science) programmes.

Scheme for 2-year M.Sc. Data Science

I (400)	**I**	**AI**	**Data Structures and Algorithms**	**Mathematics for Data Science**	**MDC-1**	**DSE-1**	**Artificial Intelligence Lab using Python**	**Data Structures and Algorithms Lab**		**22**	**44 + 4***	**Exit: PG-Dip NHEQF : 6**	**2-yr MSc NHEQF: 6.5**
	II	**Cloud Computing**	**DMBS**	**Machine Learning Techniques**	**MDC-2**	**DSE-2**	**DBMS Lab**	**ML Techniques Lab**		**22 +4***			
	Dissertation Seminar[#] ***(4 credits)***												
II (500)	**III**	**Big Data Analytics**	**Deep Learning Techniques**	**Data Visualization**	**DSE-3**	**Big Data Analytics Lab**	**Deep Learning Techniques Lab**	**Dissertation Seminar**	**Summer Training**	**24**	**44**	**1-yr MSc NHEQF : 6.5**	
	IV	**DSC-1**	**DSC-2**	**DSC-3**	**DSC-4**		**DSC-5**			**20**			
	Dissertation (20)												
	Year-long Dissertation (44)												

1. *DSE course offering would depend upon the resources.*
2. *Alternatively, department may approve suitable courses from SWAYAM.*
3. *Criterion for Semester-long Major Project: at least 7 CGPA, with no back-log.*
4. *Criterion for Year-long Major Project: at least 7.5 CGPA, with no back-log.*
5. *Request for Major Project has to be submitted to the HOD for approval well before its start.*

Semester: I

Sr. No	Course Code	Course Name	L	T	P	Credits	NEP-202 Relevancy
1.	CST415DM40	Artificial Intelligence	**3**	**1**	**0**	**4**	Artificial Intelligence skills
2.	CST417DM40	Data Structures and Algorithms	**3**	**1**	**0**	**4**	Data Structures and Algorithms skills
3.	CST419DM40	Mathematics for Data Science	**3**	**1**	**0**	**4**	Mathematics for Data Science Skills
4.		MDC-1				**4**	
5.		DSE-1	**3**	**0**	**2**	**4**	
6.	CST421DM10	Artificial Intelligence Lab using Python	**0**	**0**	**2**	**1**	Artificial Intelligence skills implementation
7.	CST423DM10	Data Structures and Algorithms Lab	**0**	**0**	**2**	**1**	Data Structures skills implementation
					Total Credits = 22		

Semester II

Sr. No.	Course Code	Course Name	L	T	P	Credits	NEP Relevancy
1.	CST456DM40	Cloud Computing	**3**	**0**	**2**	**4**	Cloud Computing skills
2.	CST458DM40	Database Management Systems	**3**	**1**	**0**	**4**	DBMS skills
3.	CST460DM40	Machine Learning Techniques	**3**	**1**	**0**	**4**	Machine Learning Skills
4.		MDC-2				**4**	
5.		DSE-2	3	0	**2**	**4**	
6.	CST462DM10	Database Management Systems Lab	**0**	**0**	**2**	**1**	DBMS skills implementation
7.	CST464DM10	Machine Learning Techniques Lab	**0**	**0**	**2**	**1**	Machine Learning Skills implementation
8	CST466DM40	Dissertation Seminar *				**4**	Research orientation
	Total Credits = 22 (22+4*= 26)						

***This course is applicable to those who opts to exit after 1st year. For this, there will be two presentations:**

- **Mid-term Presentation: Carries 40% weightage**
- **End-semester Presentation: Carries 60% weightage**

Semester: III

Sr. No	Course Code	Course Name	L	T	P	Credits	NEP2020 Relevancy
1.	CST531DM40	Big Data Analytics	**3**	**1**	**0**	**4**	Big Data Analytics skills
2.	CST533DM40	Deep Learning Techniques	**3**	**1**	**0**	**4**	Deep Learning Skills
3.	CST535DM40	Data Visualization	**3**	**0**	**2**	**4**	Data Visualization Skills
4.		DSE-3	3	0	**2**	**4**	
5.	CST537DM10	Big Data Analytics Lab	**0**	**0**	**2**	**1**	Big Data Analytics skills implementation
6.	CST539DM10	Deep Learning Techniques Lab	**0**	**0**	**2**	**1**	Deep Learning Skills implementation
7.	CST541DM40	Dissertation Seminar				**4***	Research Orientation
8	CST543DM20	Summer Training/ Internship				**2**	Real-Time Application Development
OR							
9	Year-long Dissertation					44 (of both Semesters)	Research Orientation
							Total Credits = 24

There will be two presentations:

- **Mid-term Presentation: Carries 40% weightage**
- **End-semester Presentation: Carries 60% weightage**

Semester: IV

Sr. No.	Course Code	Course Name	L	T	P	Credits	NEP Relevancy
1.	CST504SRP20	Dissertation	-	-	-	**20**	Research Orientation
OR							
2	CST516DM40	DSC-1	3	0	2	**4**	
3	CST518DM40	DSC-2	3	0	2	**4**	
4	CST522DM40	DSC-3	3	0	2	**4**	
5	CST524DM40	DSC-4	3	0	2	**4**	
6	CST526DM40	DSC-5	3	0	2	**4**	
						Total Credits = 20	

*** There will be two presentations for the dissertation:**

- **Mid-term Presentation: Carries 40% weightage**
- **End-semester Presentation: Carries 60% weightage**

DSC-Discipline Specific Course

List of Discipline Specific Electives - I

S. No	Course Code	Course Name	L	T	P	Credits	NEP Relevancy
1.	CST425DS40	Information Systems Security	3	0	2	**4**	Information Security Skills
2.	CST427DS40	Optimization Techniques	3	0	2	**4**	Optimization Techniques Skills
3.	CST429DS40	Data Warehousing and Data Mining	3	0	2	4	Data Warehousing and Data Mining skills

List of Discipline Specific Electives - II

S. No	Course Code	Course Name	L	T	P	Credits	NEP Relevancy
1.	CST468DS40	Natural Language Processing	3	0	2	**4**	Natural Language Processing Skills
2.	CST470DS40	Data Science for Internet of Things	3	0	2	**4**	Data Science for Internet of Things skills
3	CST472DS40	Time Series Analysis and Forecasting	3	0	2	4	Time Series Analysis and Forecasting skills

List of Discipline Specific Electives - III

S. No	Course Code	Course Name	L	T	P	Credits	NEP Relevancy
1.	CST445DS40	Blockchain Technology and Applications	3	0	2	**4**	Block Chain Technology Skills
2.	CST447DS40	Quantum Computing	3	0	2	**4**	Quantum Computing Skills
3.	CST449DS40	Software Engineering	3	0	2	**4**	Software Engineering Skills

11. Course-Level Learning Outcomes

Bridge Course

<table>
<tr><td>Scheme Version:
2025-onward</td><td colspan="4">Course Name:
Computer Fundamentals and C Programming</td><td colspan="2">Course Code:
CST401BC00</td></tr>
<tr><td rowspan="2">Programme:
M. Sc. Data Science</td><td rowspan="2">Total Marks: 100
CIE: 30
TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 4</td></tr>
<tr><td>3</td><td>0</td><td>2</td><td>0</td><td>Total Hours: 48(L)+32(P)=80</td></tr>
<tr><td>Semester: 1</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course aims to provide foundational knowledge of computer systems, including hardware, software, and input/output devices. It introduces students to C programming concepts such as data types, control structures, functions, and basic data structures, enabling them to write efficient and structured code.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1: Understand the evolution, classification, and components of computer systems along with their key characteristics.
CO2: Explain the working of various input, output, and storage devices used in computer systems.
CO3: Develop basic programs in C using data types, operators, control statements, and structured logic.
CO4: Implement functions, data structures, and pointer operations in C to solve programming problems effectively.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.</td></tr>
</table>

3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** **Overview of Computer System:** Evolution of Computer Systems, Generations of Computers, Parts of Computer System, Categories of Computers, Computer System Characteristics, Computer Hardware. Working of input & output devices: keyboard, mouse, trackball, pen, touch screens, scanner, digital camera, monitor, and printer. Working of storage devices: magnetic tape, magnetic disk, CD, DVD. Software- System & Application.	12
2	**[Course Outcome No(s).: 2]** **Elements of C:** Character set identifier and keywords, data type, declaration and definition. Operators: arithmetic, relational, logical, bit wise, unary, assignment and conditional operators their hierarchy arid associativity. Control statements: sequencing, selection, if and switch statement; repetition / loop statements: for, while, and do while loops; break, continue and goto statements.	11
3	**[Course Outcome No(s).: 3]** **Function:** Definition, declaration, and calling, call by value, call by reference prototype, passing parameters, actual and formal parameters, recursion.	12
4	**[Course Outcome No(s).: 4]** **Data Structures:** Arrays, structure, structure members, access to structure members union, string, data files. Pointer: declaration, operation of pointers, array to pointers, pointers to arrays.	13

Suggested Readings:

1. Norton, P., *Introduction to Computers*, Mc-Graw-Hill, 6th Ed. 2020.
2. Raja, Raman V., *Fundamental of Computers*, Prentice Hall of India,6th Ed., 2019.
3. Sanders, D. H., *Computer Today*, Mc-Graw Hill, 6th Ed., 2018.
4. Sinha, P.K. and Sinha, P., *Computer fundamentals*, BPB Publications, 8th Ed., 2020.

Semester 1

<table>
<tr><td>Scheme Version: 2025-onward</td><td colspan="5">Course Name:

Artificial Intelligence</td><td>Course Code:

CST415DM40</td></tr>
<tr><td rowspan="2">Programme: M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 4</td></tr>
<tr><td>Semester: 1</td><td>3</td><td>1</td><td>0</td><td>4</td><td>Total Hours: 48(L)+16(T)=64</td></tr>
<tr><td>Course Objectives</td><td colspan="6">The primary objective of this course is to introduce the basic principles, techniques, and applications of Artificial Intelligence. Emphasis will be placed on the teaching of these fundamentals and labs for the hands-on approach for understanding, as well as a challenging avenue for exploration and creativity.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">CO1: To introduce the fundamentals of Artificial Intelligence and its role in solving real-world data science problems.
CO2: To develop problem-solving skills using search algorithms and state-space modeling for data-driven applications.
CO3: To explore advanced AI techniques including adversarial search, constraint satisfaction, and genetic algorithms for optimization.
CO4: To understand knowledge representation and reasoning techniques, including logic-based and probabilistic approaches in uncertain environments.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** **Introduction to AI for Data Science:** Definition of Artificial Intelligence (AI) in the context of data science, Importance of AI in modern data-centric domains (healthcare, finance, e-commerce, etc.), Evolution and milestones in AI development, Overview of data-driven AI systems and their real-world impact. **Intelligent Agents and Decision Systems:** Understanding rational agents and goal-oriented behavior, Types of intelligent agents: simple reflex, model-based, goal-based, utility-based, PEAS framework: designing intelligent agents for specific tasks, Case studies: Recommendation engines, autonomous systems **AI Environments and Problem Complexity:** Classification of environments: Fully vs. partially observable, Deterministic vs. stochastic, Static vs. dynamic, Discrete vs. continuous, Relevance of these environments to data science challenges (e.g., real-time analytics, dynamic pricing).	**12**
2	**[Course Outcome No(s).: 2]** **Problem Formulation in AI Applications:** Elements of problem formulation: initial state, actions, transition model, goal test, path cost, Framing data science problems using state-space models **Uninformed Search Techniques:** Breadth-First Search (BFS), Depth-First Search (DFS), Depth-Limited Search and Iterative Deepening Search, Application of these techniques in data wrangling, preprocessing workflows **Informed Search Algorithms:** Greedy Best-First Search, A*, A* algorithm, Heuristics and cost functions, Real-world examples: route optimization, personalized recommendations **Properties and Evaluation of Search Algorithms:** Admissibility, Optimality, Monotonicity, Computational efficiency and space-time tradeoffs, Algorithm selection based on problem structure.	**12**

3	**[Course Outcome No(s).: 3]** **Adversarial Search and Game Playing:** Game theory in AI, Minimax algorithm and Alpha-Beta pruning, Applications in strategic planning. **Constraint Satisfaction Problems: Formulating CSPs:** variables, domains, and constraints, Techniques for solving CSPs (backtracking, forward checking, constraint propagation), Use cases: scheduling, timetabling, resource allocation in data workflows. **Introduction to Genetic Algorithms:** Concept of natural selection in computation, Operators: selection, crossover, mutation.	**12**
4	**[Course Outcome No(s).: 4]** **Logical Knowledge Representation:** Propositional logic: syntax, semantics, inference rules, First-order predicate logic (FOPL): syntax, semantics, unification, resolution, Rule-based systems in expert reasoning (e.g., fraud detection systems). **Structuring and Organizing Knowledge:** Semantic networks and frames, Ontologies and description logic for data categorization and knowledge graphs, Issues in knowledge representation: granularity, consistency, completeness. **Reasoning Under Uncertainty:** Bayesian reasoning and probabilistic logic, Introduction to Bayesian Networks: structure, conditional independence, use cases, Application in uncertain decision-making environments (e.g., disease prediction, market forecasting).	**12**

Suggested Readings:

1. Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach*, Pearson,4th Ed. 2020.
2. Patterson, D.W., *Introduction to Artificial Intelligence and Expert Systems*, Prentice-hall of India, 1st Ed. 2007.
3. Nilsson, N.J., *Principles of Artificial Intelligence*, Morgan Kaufmann, 1st Ed. 2014.
4. Luger, G.F. and Stubblefield, W.A., *Artificial Intelligence and The Design of Expert Systems*, Benjamin-Cummings Publishing Co. Inc., 6th Ed. 2008.

<table>
<tr><td>Scheme
Version:
2025-
onwards</td><td colspan="4">Course Name:

Data Structures and Algorithms</td><td colspan="2">Course Code:

CST417DM40</td></tr>
<tr><td rowspan="2">Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 4</td></tr>
<tr><td>3</td><td>1</td><td>0</td><td>4.0</td><td>Total Hours: 48(L)+16(T)=64</td></tr>
<tr><td>Semester: 1</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course aims to provide a solid foundation in data structures and algorithm analysis, beginning with performance evaluation techniques and basic array operations, including sorting and searching. Students will learn linear data structures such as linked lists in various forms and their memory management. The course further explores nonlinear data structures like stacks, queues, trees, and graphs, focusing on their representations, traversals, and applications. Additionally, it covers essential concepts of searching and hashing techniques, enabling efficient data management and retrieval.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1 Analyze the concepts of algorithm evaluation and find time and space complexities for searching and sorting algorithms.
CO2 Implement linear data structure such as stacks, queues, linked lists and their applications.
CO3 Implement basic operations on binary trees
CO4 Demonstrate the representation and traversal techniques of graphs and their applications</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** **Introduction to Data Structure:** Algorithm Specifications: Performance Analysis and Measurement (Time and space analysis of algorithms- Average, best- and worst-case analysis). Basic Terminology, Data structures and its classification, Representation and Analysis of Arrays, Single and Multidimensional Arrays, Address calculation, Linear search, Binary search, Traversing, Insertion & deletion in an array, Sparse matrices, Sorting techniques.	**12**
2	**[Course Outcome No(s).: 2]** **Linear Data Structure** Linked List Introduction, Representation of linked list in memory, Memory allocation – Garbage Collection, Traversing & Searching in Linked List, Insertion into linked list, Deletion from a linked list, Header Linked List, Two-way List, Input & output restricted liked list, Circular Header Linked List, Representation of polynomials using linked List.	**12**
3	**[Course Outcome No(s).: 3]** **Nonlinear Data Structure:** Stack, Array Implementation of stack, Linked Representation of Stack, Applications of stack: Conversion of Infix to Prefix and Postfix Expressions and Expression evaluation. Queue: Array and linked implementation of queues, Circular queues, D-queues and Priority Queues.	**12**
4	**[Course Outcome No(s).: 4]** **Trees**: Basic terminology, Binary Trees, algebraic Expressions, Complete Binary Tree, Extended Binary Trees, Array and Linked Representation of Binary trees, Traversing Binary trees, Threaded Binary trees, Binary Search Tree	**12**

	(BST), AVL Trees, B- trees. **Graphs**: Introduction, Sequential Representations of Graphs, Adjacency Matrices, Traversal, Connected Component and Spanning Trees, Minimum Cost Spanning Trees. **Searching & Hashing**: Sequential search, Binary search, Hash table, Hash Functions, Collision Resolution Strategies.	

Suggested Readings:

1. Lipschutz: *Data Structures (Schaum's Outline Series)*, Tata McGraw-Hill, 1st Ed. 2014.
2. Horowitz and Sahani, *"Fundamentals of data Structures"*, Galgotia Publication, 1st Ed., 1984.
3. A.M. Tenenbaum, *"Data Structures using C & C++"*, Prentice-Hall of India, 2nd Ed. 2006.
4. Trembley and Sorenson, *"Data Structures"*, TMH Publications, 1st Ed., 1984.

<table>
<tr><td>Scheme Version:
2025-onward</td><td colspan="5">Course Name:

Mathematics for Data Science</td><td>Course Code:

CST419DM40</td></tr>
<tr><td rowspan="2">Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 4</td></tr>
<tr><td>3</td><td>1</td><td>0</td><td>4.0</td><td>Total Hours: 48(L)+16(T)=64</td></tr>
<tr><td>Semester: 1</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course aims to provide students with essential mathematical tools for data science, including linear algebra concepts such as vector and matrix operations, eigenvalues, and dimensionality reduction techniques like PCA. It covers calculus fundamentals and optimization techniques vital for model training and analysis. Students will also gain a solid grounding in probability theory and statistical inference, enabling them to model uncertainty, analyze data distributions, and perform hypothesis testing and estimation for data-driven decision-making.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1 Demonstrate understanding of basic mathematical concepts in data science, relating to linear algebra, probability, and calculus.
CO2 Employ methods related to these concepts in a variety of data science applications.
CO3 Apply logical thinking to problem-solving in context.
CO4 Use appropriate technology to aid problem-solving and data analysis.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** **Linear Algebra for Data Science:** Vectors and matrices: operations, norms, dot product, matrix multiplication, Linear transformations and geometric interpretation, Eigenvalues and eigenvectors, Singular Value Decomposition, Principal Component Analysis	**12**
2	**[Course Outcome No(s).: 2]** **Calculus:** Functions, limits, and continuity, Derivatives and gradients, Multivariable calculus: partial derivatives, Jacobians, Hessians, Minimization and maximization of a function. **Optimization techniques**: Gradient descent, stochastic gradient descent, Convex and non-convex functions.	**12**
3	**[Course Outcome No(s).: 3]** **Probability Theory:** Probability spaces, conditional probability, Bayes' theorem, Random variables and distributions (discrete and continuous), Expectation, variance, covariance, moment-generating functions, **Common distributions**: Bernoulli, Binomial, Poisson, Normal, Exponential.	**12**
4	**[Course Outcome No(s).: 4]** **Statistics and Inference**: Sampling, estimators, and bias-variance tradeoff, Maximum Likelihood Estimation, Hypothesis testing and confidence intervals, t-test, chi-square test, p-values, statistical significance.	**12**

Suggested Readings:

1. Charu Aggarwal, *Linear Algebra and Optimization for Machine Learning*, 2nd Ed., PHI Learning, 2016.
2. James Stewart, *Calculus: Early Transcendentals*, Brooks/Cole, 7th Ed., 2012.
3. Norman Matloff, *Probability and Statistics for Data Science*, CRC Press, 1st Ed., 2019
4. Peter Bruce and Andrew Bruce, *Practical Statistics for Data Scientists*, O'Reilly Media, 2nd Ed. 2020.

Semester: II

<table>
<tr><td>Scheme Version:
2025-Onwards</td><td colspan="4">Course Name:

Cloud Computing</td><td colspan="2">Course Code:

CST456DM40</td></tr>
<tr><td rowspan="2">Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 5</td></tr>
<tr><td>3</td><td>0</td><td>2</td><td>4.0</td><td>Total Hours: 48(L) +32(P) = 80</td></tr>
<tr><td>Semester: 2</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course introduces students to the fundamentals of cloud computing, including its definition, key characteristics, service and deployment models, along with its benefits, challenges, and applications. It explores cloud architecture and virtualization, focusing on resource management, scalability, and the role of hypervisors and virtual machines.
Students will gain insights into major cloud service providers, cloud-based storage, databases, and networking, while also addressing critical security concerns like data privacy and compliance. The course concludes with cloud application development, migration strategies, and emerging technologies such as serverless computing, edge computing, and containerization.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1 Describe cloud service and deployment models.
CO2 Explain cloud architecture and virtualization techniques.
CO3 Evaluate cloud service providers and address security concerns.
CO4 Apply cloud programming models and discuss emerging technologies.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from</td></tr>
</table>

each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** Introduction to Cloud Computing: Definition, characteristics, service models (IaaS, PaaS, SaaS), Deployment models: public, private, hybrid, community clouds, Benefits, challenges, and applications.	**12**
2	**[Course Outcome No(s).: 2]** Cloud Architecture and Virtualization, Cloud architecture: components, resource pooling, scalability, Virtualization: types, hypervisors, virtual machines, storage virtualization.	**12**
3	**[Course Outcome No(s).: 3]** Cloud Services and Security, Cloud service providers: AWS, Azure, Google Cloud, Cloud storage, databases, networking, Security issues: data privacy, compliance, identity management.	**12**
4	**[Course Outcome No(s).: 4]** Cloud Application Development and Trends, Cloud programming models and tools, Migration to cloud, multi-cloud strategies, Emerging trends: serverless computing, edge computing, containers.	**12**

Suggested Readings:

1. Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi, *"Mastering Cloud Computing*", McGraw Hill, 3rd Ed., 2018.
2. Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, "*Cloud Computing: A Practical Approach",* McGraw Hill, 1st Ed., 2009.
3. Kai Hwang, Geoffrey C. Fox, Jack J. Dongarra, *"Distributed and Cloud Computing"*, Morgan Kaufmann, 1st Ed., 2012.
4. Thomas Erl, *"Cloud Computing: Concepts, Technology & Architecture*", Pearson, 1st Ed., 2013.

<table>
<tr><td>Scheme
Version:
2025-
onward</td><td colspan="4">Course Name:

Database Management Systems</td><td colspan="2">Course Code:

CST458DM40</td></tr>
<tr><td>Programme:

M. Sc. Data Science</td><td>Total Marks: 100

CIE: 30</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 4</td></tr>
<tr><td>Semester: 2</td><td>TEE: 70</td><td>3</td><td>1</td><td>0</td><td>4.0</td><td>Total Hours: 48(L)+16(T) =64</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course aims to provide students with a comprehensive understanding of database management systems, database design through ER diagrams, relational modeling, and integrity constraints.

Students will also gain hands-on experience with SQL and PL/SQL for querying, data manipulation, and procedural programming, including the use of triggers and exceptions. The course concludes with an introduction to NoSQL databases.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1: Understand database fundamentals, architectures, data models, and ER modeling concepts.
CO2: Apply SQL commands and normalization techniques to design and manage relational databases.
CO3: Apply normalization techniques and understand transaction management principles to ensure database consistency and efficiency.
CO4: Compare SQL and NoSQL databases and integrate DBMS with data science tools using Python or R.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** Introduction to Databases and DBMS, Characteristics of DBMS vs File Systems, Database Users and Architecture (1-tier, 2-tier, 3-tier), Data Models: Hierarchical, Network, Relational, Object-Oriented, Entity-Relationship (ER) Model: Entities, Attributes, Keys, ER Diagrams, Extended ER features: Generalization, Specialization, Aggregation, Relational Model: Relations, Schemas, Tuples, Attributes, Integrity Constraints: Domain, Key, Entity, Referential.	**12**
2	**[Course Outcome No(s).: 2]** **Structured Query Language**: DDL, DML, DCL, TCL commands, Joins, Subqueries, Views, Indexes, Grouping, Aggregation, Nested Queries.	**12**
3	**[Course Outcome No(s).: 3]** **Normalization**: 1NF, 2NF, 3NF, BCNF, 4NF, Functional Dependencies, Multi-valued Dependencies **Transaction Management**: ACID Properties	**12**
4	**[Course Outcome No(s).: 4]** **NoSQL Databases Overview:** Concept and need for NoSQL databases, Key Characteristics, Compare SQL and NoSQL databases, Type of NoSQL. **Integration of DBMS with Data Science Tools**: Python libraries (e.g., SQLAlchemy, Pandas for SQL interaction), Connecting R/Python to databases.	**12**

Suggested Readings:

1. Raghu Ramakrishnan and Johannes Gehrke, *Database Management System*, McGraw-Hill,3rd Ed., 2003.
2. Abraham Silberschatz, Henry F. Korth and Sudarshan S (2005). *Database System Concepts,* McGraw- Hill, 5th Ed., 2005.
3. Date C.J., *An Introduction to Database Systems*, 8/e, Pearson Education,8th Ed., 2003.
4. Michael Mclaughlin, *Oracle Database 11g PL/SQL Programming*, McGraw Hill, 1st Ed., 2010.
5. Shashank Tiwari, *Professional NoSQL*, John Wiley & Sons, 1st Ed., 2011.

<table>
<tr><td>Scheme Version:
2025-Onwards</td><td colspan="4">Course Name:

Machine Learning Techniques</td><td colspan="2">Course Code:

CST460DM40</td></tr>
<tr><td rowspan="2">Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 5</td></tr>
<tr><td>3</td><td>1</td><td>0</td><td>4.0</td><td>Total Hours: 48(L) +16(T) = 64</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course covers core machine learning concepts including supervised, unsupervised, and reinforcement learning, along with data preprocessing and model evaluation. It introduces regression, classification, clustering, PCA, ensemble methods, and neural networks to build and assess effective models.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">CO1: Understand the fundamentals of machine learning, including types of learning paradigms, and apply data preprocessing, feature engineering, and regression techniques for model development and evaluation.
CO2: Apply various classification algorithms and evaluate their performance using appropriate metrics such as accuracy, precision, recall, F1 score, and ROC-AUC.
CO3: Implement clustering algorithms and dimensionality reduction techniques, and assess the quality of clustering using internal evaluation metrics.
CO4: Explore and apply ensemble learning methods and basic artificial neural networks for improved model performance and generalization.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** **Introduction to Machine Learning in Data Science:** Understanding ML in the context of the data science pipeline, Overview of Supervised, Unsupervised, and Reinforcement Learning with real-world data science applications. **Data Preprocessing and Feature Engineering:** Data cleaning, handling missing values, scaling, encoding categorical variables, Feature selection methods (filter, wrapper, embedded), Feature engineering techniques to improve model performance using domain knowledge **Regression Models for Predictive Analytics:** Linear Regression (simple and multiple), Polynomial Regression, Regularization techniques: Ridge and Lasso Regression **Model Evaluation and Selection in Data Science Projects:** Understanding bias-variance trade-off, identifying overfitting and underfitting in data science workflows, Cross-validation strategies.	**12**
2	**[Course Outcome No(s).: 2]** **Classification Algorithms for Data Science:** Logistic Regression and its application in binary classification problems, k-Nearest Neighbors, Decision Trees, and Random Forests for classification tasks. **Evaluating Classification Models:** Choosing the right metric: Accuracy, Precision, Recall, F1 Score, ROC-AUC, Practical considerations in imbalanced datasets (e.g., fraud detection, medical diagnosis).	**12**
3	**[Course Outcome No(s).: 3]** **Clustering for Pattern Discovery in Data:** Algorithms: k-Means, Hierarchical Clustering, DBSCAN, Real-world clustering use cases (customer segmentation, anomaly detection) **Evaluating Clustering Results:** Silhouette Score, Davies-Bouldin Index – selecting the best clustering approach **Dimensionality Reduction in High-Dimensional Data:** Principal Component Analysis (PCA): Visualizing and simplifying complex datasets, Applications of PCA in exploratory data analysis and preprocessing	**12**

4	**[Course Outcome No(s).: 4]** **Ensemble Learning for Improved Predictive Performance**: Concepts and applications of Bagging, Boosting (e.g., AdaBoost, Gradient Boosting), and Stacking.	**12**

Suggested Readings:

1. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly, 3rd Ed. 2022.
2. Laurent Younes, *Introduction to Machine Learning*, Cambridge University Press, 2011
3. Chris Albon, *Machine Learning with Python*, O'Reilly Media, 2018.
4. Harsh Bhasin, *Machine Learning for Beginners* – 2nd Ed., 2023.

Departmental Elective Courses: Semester I

<table>
<tr><td>Scheme Version:

2025-Onwards</td><td colspan="5">Course Name:

Information Systems Security</td><td colspan="2">Course Code:

CST425DS40</td></tr>
<tr><td>Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 5</td></tr>
<tr><td>Semester: 1</td><td>3</td><td>0</td><td>2</td><td>4.0</td><td>Total Hours: 48(L) +32(P) = 80</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course provides a comprehensive understanding of information and system security, covering key concepts such as confidentiality, integrity, authentication, encryption, and threat mitigation. It explores various security threats, vulnerabilities, and protection mechanisms, including cryptographic techniques, access controls, and operating system hardening. The course also introduces cyber security practices, intrusion detection systems, and legal frameworks like the IT Act 2000 to address cybercrimes and ensure secure digital communication.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">The outcomes reflect the knowledge and skills students should acquire by the end of the course:
CO1 Understanding of fundamental information security concepts, principles, and frameworks, including confidentiality, integrity, availability, and authentication.
CO2 Knowledge of Security Threats and Vulnerabilities.
CO3 Security Mechanisms and Controls.
CO4 Risk Management and Compliance.
CO5 Security Management and Policies.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts.</td></tr>
</table>

Students are required to attempt any four sub-pats, with each carrying 3.5 marks.

3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** **Information Security: [CO1]** Need for Information Security, Privacy, Firewall, Attributes of Information Security, Aspect of Security, Confidentiality, Data Integrity, Availability, Protection Virus Security, Non-Repudiation, Authentication, Data Encryption Techniques, Hashing.	10
2	**[Course Outcome No(s).: 2]** **Security Threats, Services, Mechanism and Attacks: [CO2]** Access Control, Program Threats, Worms, Viruses, Trojan Horse, Stack and Buffer Overflow, System Threats-intruders, Communication Threats Tapping and Piracy, and Vulnerabilities, Security Attacks, Malicious Software, Authentication, Password Vulnerabilities and Attacks, Substitution, Transposition Ciphers, Symmetric key algorithms, Public key Encryption RSA, Diffie-Hellman Key Exchange, Antivirus Installation, Password Management, User Account Controls (Windows), Biometrics Techniques.	12
3	**[Course Outcome No(s).: 3]** **Physical and System Security: [CO3]** Function of Operating System, Windows Weakness, Hardening OS during Installation, Secure user Account Policy, Patching Software, Hardening Windows, Active Directory / Kerberos, Vulnerability Scanning, Manual and Automatic Hardening.	12
4	**[Course Outcome No(s).: 4 & 5]** **Cyber Security and Cyber Law: [CO4 & CO5]** Cyberspace, Cyber Crimes, Cyber Criminals, Cyber Security, Cyber Security Threats, DoS Attack, Vulnerability, Analysis, Intrusion Detection	14

	System, Web Servers and Browsers, Caching, Plug-in, ActiveX, Secure Socket Layer, Secure Electronic Transaction, Email Risks, Email Protocols, Email Security, Digital Signature, Setting up Browser Security, Email Encryption. IT Act 2000, Legal Provisions under the IT Act.	

Suggested Readings:

1. Mark Merkow & James Breithaupt, *"Information Security - Principles and Practices"*, *Pearson Education*, 1st Ed., 2007.
2. Michael. E. Whitman & Herbert J. Mattord, *"Principles of Information Security"*, Cengage Learning, 7th Ed., 2022.
3. Michael J. Palmer, *"Guide to Operating Systems Security"*. Cengage, 1st Ed., 2003.

<table>
<tr><td>Scheme
Version:
2025-
onwards</td><td colspan="4">Course Name:
Optimization Techniques</td><td colspan="2">Course Code:
CST427DS40</td></tr>
<tr><td rowspan="2">Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 5</td></tr>
<tr><td>3</td><td>0</td><td>2</td><td>4.0</td><td>Total Hours: 48(L) +32(P) = 80</td></tr>
<tr><td>Semester: 1</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course introduces the fundamental concepts and applications of Operations Research, focusing on model development, linear programming, and optimization techniques. It covers key problem-solving strategies including transportation and assignment models, sequencing, replacement strategies, and queuing theory, enabling students to analyze and optimize decision-making processes in complex systems.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1: Comprehend the techniques and applications of Engineering optimization.
CO2: Analyze characteristics of a general linear programming problem.
CO3: Apply basic concepts of mathematics to formulate an optimization problem.
CO4: Analyse various methods of solving the unconstrained minimization problem.
CO5: Analyze and appreciate variety of performance measures for various optimization problems.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** Development, definition, characteristics and phases, types of operation research models, applications; Allocation: linear programming, problem formulation, graphical solution, simplex method, artificial variables techniques, two–phase method, big-M method.	11
2	**[Course Outcome No(s).: 2]** Transportation problem: Formulation, optimal solution, unbalanced transportation problem, Degeneracy; Assignment problem, formulation, optimal solution, variants of assignment problem, traveling salesman problem.	13
3	**[Course Outcome No(s).: 3]** Sequencing: Introduction, flow, shop sequencing, n jobs through two machines, n jobs through three machines, job shop sequencing, and two jobs through "m" machines. Replacement: Introduction: Replacement of items that deteriorate with time, when money value is not counted and counted, replacement of items that fail completely, group replacement.	12
4	**[Course Outcome No(s).: 4 & 5]** Introduction, Terminology, Single Channel, Poisson arrivals and exponential service times with infinite population and finite population models, Multichannel, Poisson arrivals and exponential service times with infinite population.	12

Suggested Readings:

1 M. Natarajan, P. Balasubramani, A. Tamilarasi, *"Operations Research",* Pearson Education, 2013**.**

2 Maurice Saseini, Arhur Yaspan, Lawrence Friedman, *"Operations Research: Methods & Problems"*, 1st Ed., 1959.

3 Kalyanmoy Deb, *Optimization for Engineering Design*, PHI Publishers.

4 D. E. Goldberg, *Genetic algorithms in Search, Optimization, and Machine learning,* Addison-Wesley Publishers, 1st Ed., 1989.

<table>
<tr><td>Scheme Version:
2025-onwards</td><td colspan="5">Course Name:

Data Warehousing and Data Mining</td><td colspan="2">Course Code:

CST429DS40</td></tr>
<tr><td rowspan="2">Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 4</td></tr>
<tr><td>3</td><td>0</td><td>2</td><td>4.0</td><td>Total Hours: 48(L)+32(P) =80</td></tr>
<tr><td>Semester: 1</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course introduces the concepts of data warehousing and business analysis, focusing on data warehouse architecture, OLAP, and metadata tools. It covers data mining techniques including preprocessing, association rule mining, and classification methods such as decision trees, SVMs, and neural networks. The course also explores clustering algorithms, outlier detection, and advanced topics like spatial, text, and web mining to equip students with comprehensive knowledge for extracting insights from complex datasets.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1 Design a data mart or data warehouse for any organization
CO2 Develop skills to write queries using DMQL
CO3 Extract knowledge using data mining techniques
CO4 Adapt to new data mining tools.
CO5 Apply the techniques of clustering, classification, association finding, feature selection and visualization to real world data</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.
2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** **Data Warehousing and Business Analysis:** Data warehousing Components, Building a Data warehouse, Data Warehouse Architecture, DBMS Schemas for Decision Support, Data Extraction, Cleanup, and Transformation Tools, Metadata reporting, Query tools and Applications, Online Analytical Processing and Multidimensional Data Analysis.	**10**
2	**[Course Outcome No(s).: 2]** **Data Mining:** Data Mining Functionalities, Data Preprocessing, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization and Concept Hierarchy Generation, Architecture of A Typical Data Mining Systems-Classification of Data Mining Systems. Association Rule Mining: Efficient and Scalable Frequent Item set Mining Methods, Mining Various Kinds of Association Rules, Association Mining to Correlation Analysis, Constraint-Based Association Mining.	**12**
3	**[Course Outcome No(s).: 3]** **Classification:** Issues Regarding Classification and Prediction, Classification by Decision Tree Introduction, Bayesian Classification, Rule Based Classification, Classification by Back propagation, Support Vector Machines, Associative Classification, Lazy Learners – Other Classification Methods **Prediction**: Accuracy and Error Measures, Evaluating the Accuracy of a Classifier or Predictor, Ensemble Methods, Model Section.	**12**
4	**[Course Outcome No(s).: 4 & 5]** **Cluster Analysis and Mining:** Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods,	**14**

	Partitioning Methods, Hierarchical methods, Density-Based Methods, Grid- Based Methods, Model-Based Clustering Methods, Clustering High-Dimensional Data, Constraint-Based Cluster Analysis, Outlier Analysis. Multidimensional Analysis and Descriptive Mining of Complex Data Objects, Spatial Data Mining, Multimedia Data Mining, Text Mining, Mining the World Wide Web	

Suggested Readings:

1. Jiawei Han, Micheline Kamber and Jian Pei, *"Data Mining Concepts and Techniques",* Elsevier, 3rd Ed., 2011.
2. Alex Berson and Stephen J. Smith *"Data Warehousing, Data Mining & OLAP",* Tata McGraw – Hill Edition, 10th Reprint, 2007.
3. K.P. Soman, Shyam Diwakar and V. Ajay *"Insight into Data mining Theory and Practice",* Easter Economy Edition, Prentice Hall of India, 2006.
4. G. K. Gupta *"Introduction to Data Mining with Case Studies",* Easter Economy Edition, Prentice Hall of India, 2nd Ed., 2011.
5. Pang-Ning Tan, Michael Steinbach and Vipin Kumar *"Introduction to Data Mining",* Pearson Education, 2007.

Departmental Elective Courses: Semester II

<table>
<tr><td>Scheme
Version:
2025-
onwards</td><td colspan="4">Course Name:
Natural Language Processing</td><td colspan="2">Course Code:
CST468DS40</td></tr>
<tr><td>Programme:
M. Sc. Data Science</td><td>Total Marks: 100
CIE: 30</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 5</td></tr>
<tr><td>Semester: 2</td><td>TEE: 70</td><td>3</td><td>0</td><td>2</td><td>4.0</td><td>Total Hours: 48(L) +32(P) = 80</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course provides a foundational understanding of Natural Language Processing, covering its stages and challenges due to language ambiguity. It introduces key concepts in information theory, character and sentence segmentation, morphological analysis using finite state techniques, and statistical methods such as n-gram models and hypothesis testing. Learners also explore word sense disambiguation through supervised and dictionary-based approaches, enabling effective language modeling and analysis.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">Upon successful completion of the course students will be able to:
CO1 Will be able to understand the wide spectrum of problem statements, tasks, and solution approaches within NLP
CO2 Will be able to implement and evaluate different NLP applications and apply machine learning and deep learning methods for this process.
CO3 Evaluate various algorithms and approaches for the given task, dataset, and stage of the NLP product.
CO4 Understand best practices, opportunities, and the roadmap for NLP from a business and product leader’s perspective.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
1. Question Paper will consist of five questions.</td></tr>
</table>

2. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
3. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** Introduction to NLP - Various stages of NLP –The Ambiguity of Language: Why NLP Is Difficult Parts of Speech: Nouns and Pronouns, Words: Determiners and adjectives, verbs, Phrase Structure. Statistics Essential Information Theory: Entropy, perplexity, The relation to language, Cross entropy.	**12**
2	**[Course Outcome No(s).: 2]** Character Encoding, Word Segmentation, Sentence Segmentation, Introduction to Corpora, Corpora Analysis. Inflectional and Derivation Morphology, Morphological analysis and generation using Finite State Automata and Finite State transducer.	**12**
3	**[Course Outcome No(s).: 3]** Words: Collocations- Frequency-Mean and Variance – Hypothesis testing: T-test, Hypothesis testing of differences, Pearson's chi-square test, Likelihood ratios. Statistical Inference: n -gram Models over Sparse Data: Bins: Forming Equivalence Classes- N gram model - Statistical Estimators- Combining Estimators.	**12**
4	**[Course Outcome No(s).: 4]** Methodological Preliminaries, Supervised Disambiguation Bayesian classification, an information theoretic approach Dictionary-Based Disambiguation: Disambiguation based o sense, Thesaurus based disambiguation, Disambiguatio based	**12**

	on translations in a second-language corpus.	

Suggested Readings:

1. Nitin Indurkhya, Fred J. Damerau "*Handbook of Natural Language Processing*", 2nd Ed., CRC Press, 2010.
2. James Allen "*Natural Language Understanding*", Pearson Publication, 8th Ed., 2012.
3. Chris Manning and Hinrich Schütze, *"Foundations of Statistical Natural Language Processing"*, MIT Press, 2nd Ed., 2003.
4. Hobson lane, Cole Howard, Hannes Hapke, *"Natural Language Processing in Action*" Manning Publications, 1st Ed., 2019.

<table>
<tr><td>Scheme Version:
2025-onwards</td><td colspan="4">Course Name:

Data Science for Internet of Things</td><td colspan="2">Course Code:

CST470DS40</td></tr>
<tr><td>Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 5</td></tr>
<tr><td>Semester: 2</td><td>3</td><td>0</td><td>2</td><td>4.0</td><td>Total Hours 48(L) +32(P) = 80</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course provides a comprehensive introduction to the intersection of Data Science and IoT. Students will learn to collect, preprocess, and analyze time-series data from sensors, and apply machine learning techniques such as anomaly detection and predictive maintenance tailored to IoT applications. The course also explores cloud platforms and big data tools for IoT, along with essential topics in data privacy and security, enabling students to build scalable, intelligent, and secure IoT solutions.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">After completion of this course, the students will be able to
CO1: Understand the architecture of IoT systems and analyze the role of data science in processing diverse IoT devices.
CO2: Apply data acquisition and preprocessing techniques to IoT data collected from sensors and edge devices.
CO3: Implement machine learning models for IoT applications such as anomaly detection, predictive maintenance, and device profiling, with an emphasis on lightweight models for edge devices.
CO4: Utilize cloud and big data platforms like AWS IoT Core, Azure IoT Hub, and Apache Kafka to manage and process large-scale IoT data streams efficiently.
CO5: Demonstrate awareness of data privacy and security issues in IoT systems, including GDPR compliance, data anonymization, and mitigation strategies for common IoT security threats.</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
4. Question Paper will consist of five questions.
5. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
6. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** Overview of IoT systems and architecture, Role of data science in IoT ecosystems, Sources and types of IoT data (sensor, log, streaming), IoT communication protocols (MQTT, CoAP, HTTP), Challenges in IoT data analysis: volume, variety, velocity, veracity	**11**
2	**[Course Outcome No(s).: 2]** **Data Acquisition and Preprocessing**: Data collection from sensors, edge devices, and gateways, Data wrangling: cleaning, filtering, aggregation, Handling missing data and noise, Data normalization and transformation, Time-series data and event-based data preprocessing, Descriptive statistics and feature extraction, Correlation, anomaly, and trend detection, Case study: EDA on environmental or industrial IoT dataset	**13**
3	**[Course Outcome No(s).: 3]** **Machine Learning for IoT Data**: Supervised and unsupervised learning for IoT, Anomaly detection in sensor data, Predictive maintenance using regression and classification, Clustering for device behavior profiling. Edge AI: lightweight models for constrained devices	**12**
4	**[Course Outcome No(s).: 4 & 5]** **Cloud and Big Data Platforms for IoT**: Cloud platforms: AWS IoT Core, Azure IoT Hub, Google Cloud IoT, Stream processing tools: Apache Kafka, Spark Streaming, AWS Kinesis, Data lakes and storage architectures for IoT, Integration of ML pipelines in the cloud, Data privacy issues in IoT (GDPR, consent, anonymization), IoT security threats and mitigation.	**10**

Suggested Readings:

1. Rachel Schutt & Cathy O'Neil, *"Doing Data Science"* O' Reilly, 1st Ed., 2013.

2. Robert Stackowiak, Art Licht, Venu Mantha, Louis Nagode, *Big Data and The Internet of Things: Enterprise Information Architecture for A New Age,* APress, 1st Ed., 2015
3. Peter Waher, *Learning Internet of Things,* Packt Publishing, 1st Ed., 2015
4. Dirk Slama, Frank Puhlmann, Jim Morrish, *Enterprise IOT*, O'Reilly, 1st Ed., 2015
5. Adrian McEwen, Hakim Cassimally, *Designing the Internet of Things*, John Wiley and Sons, 1st Ed., 2014

<table>
<tr><td>Scheme Version:

2025-onwards</td><td colspan="5">Course Name:

Time Series Analysis and Forecasting</td><td colspan="2">Course Code:

CST472DS40</td></tr>
<tr><td>Programme:

M. Sc. Data Science</td><td rowspan="2">Total Marks: 100

CIE: 30

TEE: 70</td><td>L</td><td>T</td><td>P</td><td>Credits</td><td>Contact Hours per Week: 5</td></tr>
<tr><td>Semester: 3</td><td>3</td><td>0</td><td>2</td><td>4.0</td><td>Total Hours: 48(L)+32(P)= 80</td></tr>
<tr><td>Course Objectives</td><td colspan="6">This course provides a comprehensive understanding of time series analysis, focusing on the characteristics and components of stochastic processes and various types of stationarity. It covers foundational models like MA, AR, ARMA, and ARIMA, along with techniques for estimating model coefficients and assessing model fit using criteria like AIC and BIC. Students will also explore methods for detecting unit roots and addressing spurious trends through statistical tests such as ADF and Phillips-Perron.</td></tr>
<tr><td>Course Outcomes:</td><td colspan="6">By the end of the course, students will be able to
CO1 Pre-process time series data and identify relevant patterns such as trends and seasonality
CO2 Select appropriate models for data modeling
CO3 Implement these models in Python and generate predictions
CO4 Evaluate the forecasting performance of the chosen models and understand their limitations</td></tr>
<tr><td colspan="7">COURSE SYLLABUS</td></tr>
<tr><td colspan="7">Instructions for the paper-setter:
Maximum Marks = 70 Time: 3 Hours
Weightage per Unit = 14 marks
7. Question Paper will consist of five questions.
8. Question No. 1 will cover the entire syllabus and will consist of seven sub-parts. Students are required to attempt any four sub-pats, with each carrying 3.5 marks.
9. Questions 2 to 5 will be set from all four units of the syllabus, with one question from each unit. Each question will consist of three sub-parts, out of which students must attempt any two. Each sub-part will carry 7 marks.</td></tr>
</table>

Unit No.	Content of Each Unit	Hours of Each Unit
1	**[Course Outcome No(s).: 1]** Time series characteristics, types, components, Stochastic process and its main characteristics. Time series as a discrete stochastic process. Stationarity. Main characteristics of stochastic processes (means, autocovariation and autocorrelation functions). Stationary stochastic processes. Stationarity as the main characteristic of stochastic component of time series. Types of stationarity: Strict, Weak, Wold decomposition. Lag operator, Time Series vs Cross-Sectional Data, Applications of Time Series Forecasting.	**12**
2	**[Course Outcome No(s).: 2]** Moving average models MA(q). Condition of invertibility. Autoregressive models AR(p). Yull-Worker equations. Stationarity conditions. Autoregressive-moving average models ARMA (p,q). Non-seasonal and seasonal ARIMA models	**12**
3	**[Course Outcome No(s).: 3]** Coefficients estimation in autoregressive models. Coefficient estimation in ARMA (p) processes. Quality of adjustment of time series models. AIC information criterion. BIC information criterion. "Portmonto"-statistics. Box-Jenkins methodology to identification of stationary time series models.	**12**
4	**[Course Outcome No(s).: 4]** The unit root problem. Spurious trends and regressions. Unit root tests (Dickey-Fuller). ADF test and the choice of the number of lags. Other unit root tests: Phillips-Perron (PP) Test, Zivot-Andrews Test, Random **Walk,** Detrending and Differencing.	**12**

Suggested Readings:

1. Econometric Views 4.0 User's Guide. *Quantitative Micro Software*, LLC, 2007.
2. Banerjee, A., J.J. Dolado, and D.V. Hendry. *Co-Integration, Error Correction, and*

Econometric Analysis of Non-Stationary Data. Oxford University Press, 1st Ed., 1993

3. Maddala, G.S. And Kim In-Moo. *Unit Roots, Cointegration, and Structural Change.* Cambridge University Press, 1st Ed., 1998
4. P. J. Brockwell, R. A. Davis, *Introduction to Time Series and Forecasting*. Springer, 3rd Ed., 2016.
5. J. Johnston, J. DiNardo. *Econometric Methods*. McGraw-Hill, 4th Ed., 1997.

12 **Teaching-Learning Process**

- Lectures
- Discussions
- Simulations
- Role Playing
- Participative Learning
- Interactive Sessions
- Seminars
- Research-based Learning/Dissertation or Project Work
- Technology-embedded Learning

13. Implementation of Blended Learning

Blended Learning is a pedagogical approach that combines face to-face classroom methods with computer-mediated activities in the process of teaching and learning. It implies nice blend of face-to-face and online activities to make the learning processes more interesting and engaging. It focuses on integration of traditional classroom activities and innovative ICT-enabled strategies. It emphasizes student-centric learning environment where the teacher is the facilitator for productive and measurable learning outcomes. It optimizes and compliments the face to face learning, giving ample freedom and flexibility to the students and teachers to access and explore the wide range of open-access sources such as video lectures, podcasts, recordings and articles through digital platforms. It gives freedom and autonomy to the teachers in selection of appropriate digital platforms, resources and time-slots to complement and supplement face to face learning. The Blended Learning doesn't undermine the role of the teacher, rather it gives him/her an opportunity to explore the unexplored in accordance with the requirements of the curriculum.

Key features of Blended Learning

- **Student-Centric Pedagogical Approach** focusing on flexibility in timing, quality content, needs and interests of students and freedom to study through the mode of his/her choice;
- Freedom to Select variety of mediums and techniques;
- Increased student engagement in learning;
- Enhanced teacher and student interaction;
- Improved student learning outcomes;
- More flexible teaching and learning environment;
- More responsive for self and continuous learning;
- Better opportunities for experiential learning;
- Increased learning skills;
- Greater access to information, improved satisfaction and learning outcomes.

Note: Resolution no (c) as per minutes circulated by VC office**:** It was resolved that Blended Learning with 40% component of online teaching and 60% face to face classes for each programme, be adopted.

14. Assessment and Evaluation

- Continuous Comprehensive Evaluation at regular after achievement of each Course-level learning outcome
- Formative Assessment on the basis of activities of a learner throughout the programme instead of one-time assessment
- Oral Examinations to test presentation and communication skills
- Open Book Examination for better understanding and application of the knowledge acquired
- Group Examinations on Problem solving exercises
- Seminar Presentations
- Review of Literature
- Collaborative Assignments

15. Keywords

a. LOCF
b. NEP-2020
c. Blended Learning
d. Face to face (F to F) Learning
e. Programme Outcomes
f. Programme Specific Outcomes
g. Course-level Learning Outcomes
h. Postgraduate Attributes
i. Learning Outcome Index
j. Formative Assessment and Evaluation
k. Comprehensive and Continuous Evaluation

16 References

l. National Education Policy-2020. https://www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf

m. The draft subject specific LOCF templates available on UGC website. https://www.ugc.ac.in/ugc_notices.aspx?id=MjY5OQ==

n. Draft Blended Mode of Teaching and Learning: Concept Note available on UGC website. https://www.ugc.ac.in/pdfnews/6100340_Concept-Note-Blended-Mode-of-Teaching-and-Learning.pdf

17. Appendices

- **Syllabi & Scheme**