COURSE STRUCTURE AND SYLLABI

M.Tech in Data Science

2025-26 Batch

SCHOOL OF ENGINEERING AND TECHNOLOGY

CENTURION UNIVERSITY OF TECHNOLOGY & MANAGEMENT

Odisha-761211, India

Web Site: - www.cutm.ac.in

CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT, ODISHA

CERTIFICATE

Centurion UNIVERSITY

Shaping Lives...
Empowering Communities...

This is to certify that the syllabus of the Programme M.Tech in Data Science of the School of Engineering and Technology is approved in the 15th Academic Council Meeting held on 22nd November 2025.

Dean
School of Engineering and Technology,
CUTM, Odisha

Pro Vice Chancellor
CUTM, Odisha

Centurion University of Technology and Management Odisha

M.Tech. in Data Science

COURSE STRUCTURE & SYLLABUS

School of Engineering & Technology

2025

Index

Programme Objectives; Job/Higher studies/Entrepreneurship

POs: Engineering Graduates will be able to;

1. **PO1:** Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **PO2:** Problem Analysis: Identify, formulate, review research literature, and analyze complex engineering problems to reach substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **PO3:** Design/Development of Solutions: Design solutions for complex engineering problems and design system components or processes that meet specified needs with appropriate consideration for public health and safety, and cultural, societal, and environmental considerations.
4. **PO4:** Conduct Investigations of Complex Problems: Use research-based knowledge and research methods including the design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **PO5:** Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **PO6:** The Engineer and Society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to professional engineering practice.
7. **PO7:** Environment and Sustainability: Understand the impact of professional engineering solutions in societal and environmental contexts, and demonstrate knowledge of, and need for, sustainable development.
8. **PO8:** Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **PO9:** Individual and Team Work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **PO10:** Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **PO11:** Project Management and Finance: Demonstrate knowledge and understanding of engineering 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.
12. **PO12:** Life-long Learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

PEOs/PSOs

1. **PEO1:** Graduates will excel in professional careers by applying advanced knowledge of data science, machine learning, and computational techniques to develop scalable and intelligent solutions across diverse domains.
2. **PEO2:** Graduates will demonstrate strong analytical, research, and problem-solving capabilities, enabling them to contribute innovatively to scientific discovery, industrial advancement, and interdisciplinary projects.
3. **PEO3:** Graduates will exhibit leadership, ethical responsibility, effective communication, and adaptability to emerging technologies, fostering continuous learning and societal impact.
1. **PSO1:** Graduates will be able to design, implement, and optimize advanced data-driven models using statistical methods, machine learning algorithms, and big data frameworks.
2. **PSO2:** Graduates will effectively manage, process, analyze, and visualize complex datasets to generate meaningful insights for decision-making in real-world applications.
3. **PSO3:** Graduates will apply state-of-the-art tools, cloud technologies, and domain-specific knowledge to develop scalable, secure, and intelligent data science solutions that address industry and research challenges

Course Outcomes	Attributes
CO1	Knowledge
CO2	Analytical skill and Critical Thinking
CO3	Problem Solving and Decision taking ability
CO4	Use of Tool, Design and Development (Hands-on/ Technical skill)
CO5	Research, Ethics & Team work

Master of Technology(M.Tech in Data Science)	
Programme Structure – Post-Graduate Study	
Type of Course	2 Years
Major (Core) Courses	40
Elective Core/ Domain	4
Skill + AEC	4
Research Project	32
TOTAL	80

Course Structure

SEMESTER - I					
Sl No	**Code**	**Subject name**	**Credits**	**T+P+J**	**NcRF Level**
1	MTDS1101	Introduction to Data Science	4	2+2+0	6.5
2	MTDS1102	Mathematical Foundation for Data Science	4	2+2+0	6.5
3	MTDS1103	Machine Learning Using Python	4	0+4+0	6.5
4	MTDS1104	Data Mining and Data Warehousing	4	2+2+0	6.5
5	MTDS1105	Statistical Methods	4	2+2+0	6.5
6	CUTM3191	Generative AI and Prompt Engineering	4	0+2+2	6.5
		TOTAL	**24**		
SEMESTER - II					
1	MTDS1201	Design and Analysis of Algorithms	4	2+2+0	6.5
2	MTDS1202	Big Data Systems	4	2+2+0	6.5
3	MTDS1203	Digital Image Processing	4	0+4+0	6.5
4	MTDS1204	Information Retrieval	4	2+2+0	6.5
5	MTDS1205	Computational Intelligence	4	2+2+0	6.5
6	CUTM2378	Research Methodology and IPR	4	2+0+2	6.5
		TOTAL	**24**		
SEMESTER - III					
1	MTIP 2101	Industry Internship & Project I	16	0+0+16	7
		TOTAL	**16**		
SEMESTER - IV					
1	MTIP 2201	Industry Internship & Project II	18	0+0+18	7
		TOTAL	**18**		

Introduction to Data Science (60 Hours)

Course	Code	T-P-Pr	Credit
Introduction to Data Science	MTDS1101	2-2-0	4

Course Objectives:

- To introduce fundamental concepts of data science using Python, SQL/SPARQL, and Map-Reduce techniques for handling small and big data.
- To develop skills in data collection, preparation, querying, analytics, visualization, and parallel computing.
- To provide in-depth understanding of advanced data analytics including NLP, graph analytics, knowledge extraction, and knowledge bases.

Course Outcomes:

At the end of the course, students will be able to:

- **CO1:** Apply Python, SQL/SPARQL, and Map-Reduce techniques for effective data manipulation and analysis.
- **CO2:** Perform data collection, preparation, querying, and implement pattern mining, classification, clustering, and visualization techniques.
- **CO3:** Demonstrate advanced analytics skills including NLP, knowledge extraction, graph analytics, and querying over knowledge bases.
- **CO4:** Utilize parallel and distributed computing platforms for scalable data analysis.
- **CO5:** Apply data science methods to real-world domains such as business intelligence, social media analytics, biomedical informatics, computational ecology, and legal

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3												3	2	
CO2		3	2										3	3	
CO3			3	2									3	3	2
CO4	2		3		2								2	3	
CO5		2							3				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

This course will give an introduction to the basic data science techniques including programming in Python, SQL/SPARQL and Map-Reduce for small and big data manipulation and analytics.

Module II (15 Hours):

Data collection, data preparation, data querying, data analytics including pattern mining, classification, clustering, data visualization, and parallel computing platforms.

Module III (15 Hours):

Advanced data analytics including NLP, knowledge extraction, graph analytics, graph querying, knowledge bases and crowd sourcing.

Module IV (15 Hours):

Introduce key application areas of data science including business intelligence, social media, biomedical informatics, computational ecology and e-discovery.

Text Books:

1. Data Science from Scratch (DSS), Joel Grus, O'Reilly Media Inc., http://shop.oreilly.com/product/0636920033400.
2. Python for Data Analysis (PDA), Wes McKinney, O'Reilly Media Inc., http://proquest.safaribooksonline.com/9781449323592
3. Mining of massive datasets (MMD), A. Rajaraman and J.D. Ullman, Cambridge University Press, 2011. ISBN-10: 1107015359, ISBN-13: 978-1107015357, http://www.mmds.org/ (public online access).
4. Natural Language Processing with Python (NLTK Book): http://www.nltk.org/book/ (public online access).

Reference Books:

1. Learning scikit-learn: Machine Learning in Python (MLP), Guillermo Moncecchi and Raul Garreta
2. Doing Data Science (DDS), Cathy O'Neil and Rachel Schutt, O'Reilly Media Inc., http://proquest.safaribooksonline.com/9781449363871

Mathematical Foundation for Data Science (60 Hours)

Subject Name	Code	T-P-P	(Credit)
Mathematical Foundation for Data Science	MTDS1102	2-2-0	4

Course Objectives:

- To introduce the mathematical foundations essential for data science, including linear algebra, probability, statistics, and optimization.
- To develop structured thinking and problem-solving abilities for addressing data science problems.
- To build foundational understanding of optimization and its role in machine learning and data science applications.

Course Outcomes (COs):

At the end of the course, students will be able to:

- **CO1:** Demonstrate understanding of basic data science concepts, problem typologies, and structured approaches for mathematical modeling.
- **CO2:** Apply linear algebra concepts such as matrices, eigenvalues, projections, and hyperplanes to formulate and analyze data-oriented problems.
- **CO3:** Utilize probability, statistics, and random process concepts for data modeling, estimation, and hypothesis testing.
- **CO4:** Apply unconstrained and constrained optimization techniques, gradient-based methods, and least-squares solutions to data science and machine learning problems.
- **CO5:** Understand and apply foundational data science methods such as linear regression and linear classification.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3												2	2	
CO2		3	3										3	2	
CO3	2	3		2									2	3	2
CO4			3		3								3	3	
CO5	2								2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

Basics of Data Science: Introduction; Typology of problems; Importance of linear algebra, statistics and optimization from a data science perspective; structured thinking for solving data science problems.

Module II (15 Hours):

Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigen values and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplane; half-planes.

Module III (15 Hours):

Probability, Statistics and Random Processes: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate and multivariate); Expectations and moments; Covariance and correlation; Statistics and sampling distributions; Hypothesis testing of means, proportions, variances and correlations; Confidence (statistical) intervals; Correlation functions; White-noise process.

Module IV (15 Hours):

Optimization: Unconstrained optimization; Necessary and sufficiency conditions for optima; Gradient descent methods; Constrained optimization, KKT conditions; Introduction to non-gradient techniques; Introduction to least squares optimization; Optimization view of machine learning.5. Introduction to Data Science Methods: Linear regression as an exemplar function approximation problem; linear classification problems.

Text Books:

1. G. Strang (2016). Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition, USA.
2. Bendat, J. S. and A. G. Piersol (2010). Random Data: Analysis and Measurement Procedures. 4th Edition. John Wiley & Sons, Inc., NY, USA:
3. Montgomery, D. C. and G. C. Runger (2011). Applied Statistics and Probability for Engineers. 5th Edition. John Wiley & Sons, Inc., NY, USA:
4. David G. Luenberger (1969). Optimization by Vector Space Methods, John Wiley & Sons (NY)

Reference Books:

Cathy O’Neil and Rachel Schutt (2013). Doing Data Science, O’Reilly Media

Machine Learning using Python (60 Hours)

Subject Name	Code	T-P-P	(Credit)
Machine Learning using Python	MTDS1103	0-4-0	4

Course Objectives:

- Understand the fundamental concepts of machine learning, including supervised and unsupervised learning, neural networks, probabilistic models, and optimization techniques.
- Apply various machine learning algorithms such as decision trees, clustering, dimensionality reduction, and evolutionary algorithms to analyze and solve real-world data problems.
- Develop practical skills in implementing machine learning models using Python libraries such as NumPy, Matplotlib, and related tools for data processing, visualization, and model evaluation.

Course Outcomes (COs):

At the end of the course, students will be able to:

- **CO1:** Understand different learning paradigms and implement basic supervised learning models.
- **CO2:** Construct decision trees, probabilistic models, Gaussian mixtures, and nearest neighbor classifiers.
- **CO3:** Apply clustering, dimensionality reduction, optimization, and evolutionary computing techniques.
- **CO4:** Analyze Bayesian networks, Markov models, and tracking methods.
- **CO5:** Implement machine learning algorithms using Python libraries.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3												2	2	
CO2		3	3										3	2	
CO3	2	3		2									2	3	2
CO4			3		3								3	3	
CO5	2								2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

Learning - Types of machine learning - Supervised learning - The brain and the neurons, Linear Discriminants -Perceptron - Linear Separability -Linear Regression - Multilayer perceptron – Examples of using MLP - Back propagation of error.

Module II (15 Hours):

Decision trees- Constructing decision trees-Classification of regression trees- Regression example -Probability and Learning: Turning data into probabilities - Some basic statistics - Gaussian mixture models - Nearest Neighbor methods.

Module III (15 Hours):

The k-Means algorithm - Vector Quantization's - Linear Discriminant Analysis - Principal component analysis - Factor Analysis - Independent component analysis - Locally Linear embedding – Isomap - Least squares optimization - Simulated annealing. The Genetic algorithm - Genetic operators - Genetic programming - Combining sampling with genetic programming - Markov Decision Process - Markov Chain Monte Carlo methods: sampling - Monte carlo - Proposal distribution.

Module IV (15 Hours):

Bayesian Networks - Markov Random Fields – Hidden Markov Models -Tracking methods. Python: Installation –Python for MAT LAB and R users-Code Basics –Using NumPy and MatPlotLiB.

Text Books:

1. Kevin P. Murphy, "Machine Learning – A probabilistic Perspective", MIT Press, 2016.
2. Randal S, "Python Machine Learning, PACKT Publishing, 2016.

Reference Books:

1. Ethem Alpaydin, "Machine Learning: The New AI", MIT Press, 2016.
2. Shai Shalev-Shwartz, Shai Ben-David, "Understanding Machine Learning: From Theory to Algorithms", Cambridge University Press, 2014.
3. Sebastian Raschka, "Python Machine Learning", Packt Publishing Ltd, 2015.

Data Mining and Data Warehousing (60 Hours)

Subject Name	Code	T-P-P	(Credit)
Data Mining and Data Warehousing	MTDS1104	2-2-0	4

Course Objectives:

- To provide a comprehensive understanding of data mining concepts, techniques, and applications for knowledge discovery.
- To equip students with skills in classification, clustering, pattern mining, and advanced analytical techniques.
- To develop the ability to analyze complex, high-dimensional, temporal, spatial, and web data.
- To introduce the fundamentals of data warehousing and its integration with data mining techniques for real-world applications.

Course Outcomes (COs):

At the end of the course, students will be able to:

- **CO1:** Understand and apply data mining tasks including frequent pattern mining, classification, clustering, correlation analysis, and outlier detection.
- **CO2:** Implement classification techniques using backpropagation, SVMs, frequent pattern-based methods, and soft computing approaches.
- **CO3:** Apply density-based, grid-based and high-dimensional clustering algorithms to various types of datasets.
- **CO4:** Analyze and perform web mining, text mining, temporal data mining, and spatial data mining tasks.
- **CO5:** Demonstrate the ability to apply data mining methods for solving real-world problems across domains.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3												3	2	
CO2		3	2										3	3	
CO3	2	3		2									3	3	2
CO4			3		3								2	3	
CO5		2							2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

Data mining tasks – mining frequent patterns, associations and correlations, classification and regression for predictive analysis, cluster analysis , outlier analysis; advanced pattern mining in multilevel, multidimensional space – mining multilevel associations, mining multidimensional associations, mining quantitative association rules, mining rare patterns and negative patterns.

Module II (15 Hours):

Classification by back propagation, support vector machines, classification using frequent patterns, other classification methods – genetic algorithms, roughest approach, fuzz>set approach.

Module III (15 Hours):

Density – based methods –DBSCAN, OPTICS, DENCLUE; Grid-Based methods – STING, CLIQUE; Exception – maximization algorithm; clustering High- Dimensional Data; Clustering Graph and Network Data.

Module IV (15 Hours):

Introduction, web mining, web content mining, web structure mining, we usage mining, Text mining – unstructured text, episode rule discovery for texts, hierarchy of categories, text clustering.
Temporal and Spatial Data Mining: Introduction; Temporal Data Mining – Temporal Association Rules, Sequence Mining, GSP algorithm, SPADE, SPIRIT Episode Discovery, Time Series Analysis, Spatial Mining – Spatial Mining Tasks, Spatial Clustering. Data Mining Applications.

Text Books:

1. Data Mining Concepts and Techniques, Jiawei Hang Micheline Kamber, Jian pei, Morgan Kaufmannn.
2. Data Mining Techniques – Arun K pujari, Universities Press.

Reference Books:

Introduction to Data Mining – Pang-Ning Tan, Vipin kumar, Michael Steinbach, Pearson.
Data Mining Principles & Applications – T.V Sveresh Kumar, B.Esware Reddy, Jagadish S Kalimani, Elsevier.

Statistical Methods (60 Hours)

Subject Name	Code	T-P-P	(Credit)
Statistical Methods	MTDS1105	2-2-0	4

Course Objectives:

- To develop a strong foundation in basic and advanced statistical concepts for data analysis.
- To introduce probability theory, probability distributions, and multivariate statistical techniques.
- To equip students with skills in regression modeling, parameter estimation, and significance testing.
- To familiarize students with pattern analysis and geostatistical methods relevant to engineering and data science applications.

Course Outcomes (COs):

At the end of the course, students will be able to:

- **CO1:** Understand and apply basic statistical concepts including measures of central tendency, dispersion, correlation, and regression.
- **CO2:** Apply probability theory, probability distributions, and Bayesian concepts to model and analyze data.
- **CO3:** Perform multivariate analysis including multivariate data summaries, covariance matrices, and multiple regression modeling.
- **CO4:** Implement and analyze advanced statistical techniques including autocorrelation, semivariogram analysis, Kriging, and pattern analysis.
- **CO5:** Interpret statistical results to support decision making in scientific, engineering, and data-driven applications.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3												2	2	
CO2		3	2										3	3	
CO3	2	3		2									2	3	2
CO4			3		3								3	3	
CO5	2								2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

Basic Statistics: Sources of Data, Organization of Data, The Histogram, Measures of central tendency, Mean Deviation, Standard Deviation, Correlation, Coefficient of correlation, Rank correlation, Regression.

Module II (15 Hours):

Probability: equally likely, mutually exclusive events, definitions of probability, additions & multiplication theorems of probability and problems based on them. Bayesian approach, distributions; Poisson, normal, Erlang, Gamma and Weibull probability distributions. Multivariate Data:Random Vectors and Matrices, sample estimate of centroid, standard deviation, SSCP, dispersion, variance, covariance, correlation matrices.

Module III (15 Hours):

Multivariate Regression Models, Multiple linear Regression: Multiple parameter estimation by method of least squares, tests of significance use of dummy variables, problems associated with multi colinearity, heteroscadasticity.

Module IV (15 Hours):

Pattern Analysis, Measures of Arrangements & dispersion, Auto Correlation, Semiveriogram, Kriging

Text Books:

1. Gupta, S.C. and Kapoor, V.K., “Fundamentals of Mathematics Statistics”, Sultan Chand and Sons, 2001.
2. Johnson, R.J., “Miller and Freund’s Probability and Statistics for Engineers” 6th Edition, Prentice Hall of India, 2002.

Reference Books:

1. Jay L. Devore, “Probability and statistics for Engineering and the Sciences”, Thomson and Duxbbury, 2002.
2. Sarma, D.D. “Geostatistics with Applications in Earth Sciences”, Capital Publishing Company, 2002.
3. Cooley W. W and Lohnes P.R .- Multivariate Data Analysis, John Wiley and Sons,1971.

Generative AI and Prompt Engineering (60 Hours)

Subject Name	Code	T-P-P	(Credit)
Generative AI and Prompt Engineering	CUTM3191	0-2-2	4

Course Objectives:

- To introduce the foundations, evolution, and applications of Generative AI, GPT models, and large-scale foundation models.
- To develop strong skills in prompt engineering, optimization strategies, and evaluation methodologies for AI-generated outputs.
- To provide hands-on experience with modern AI development frameworks, APIs, and multimodal generative technologies.

Course Outcomes (COs):

At the end of the course, students will be able to:

- **CO1:** Understand the evolution, architecture, and capabilities of Generative AI models including GANs, Transformers, LLMs, and GPT variants.
- **CO2:** Apply prompt engineering principles, prompting techniques, and evaluation metrics to optimize AI-generated responses.
- **CO3:** Design and implement RAG pipelines, vector databases, and multimodal AI applications for real-world tasks.
- **CO4:** Build agentic AI systems using frameworks such as LangChain Agents, CrewAI, AutoGen, and workflow automation tools.
- **CO5:** Evaluate generative AI models for ethics, bias, cultural sensitivity, transparency, and responsible usage.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2			3								3	2	
CO2		3	3										3	3	
CO3	2	3		3									3	3	2
CO4			3		3								2	3	
CO5		2							2				2	2	3

***High-3, Medium-2, Low-1**

Module 1: Generative AI, GPT and Foundation Models (15 Hours)

Theory:

Evolution of Generative AI. Mathematical and computational foundations of generative modelling. Generative Adversarial Networks (GANs). Large Language Models (LLMs) and architectures (Transformers, BERT, GPT), Applications of Generative AI across education, healthcare, business, and creative fields.

GPT (Generative Pre-trained Transformer): Architecture and functionality. Pre- training, fine-tuning, and adaptation for domain-specific tasks. Evolution of GPT models: GPT-2, GPT-3, GPT-4, and beyond. ChatGPT and conversational AI systems.

Practice:

- Design a prompt for an AI-based educational tool.
- Evaluate generated text for coherence and diversity.
- Create a basic prompt for a chatbot interaction.
- Develop a simple text-generation model using OpenAI API.
- Optimize a prompt for better user engagement using GPT models.
- Evaluate the response quality of different AI models (GPT, Gemini, Claude).
- Analyse the effectiveness of fine-tuning for domain adaptation.
- Design a prompt for a language translation or summarization task.

Module 2: Prompt Engineering and Development Frameworks (15 Hours)

Theory:

Concept of Prompt engineering and optimization. Types of prompts: Zero-shot, One- shot, Few-shot Prompting. Advanced prompting techniques: Chain-of-Thought (CoT), Tree-of-Thought (ToT), Self-Consistency, Step-back, Least-to-Most, and Adversarial Prompting. Evaluation metrics: BLEU, ROUGE, and hallucination detection. Hands-on with OpenAI, Gemini, Hugging Face, Claude APIs. LangChain Framework: Components, architecture, and workflow.

Practice:

- Analyze the effectiveness of prompts in a given scenario.
- Design prompts for specific tasks such as summarization or question answering.
- Develop a prompt-based recommendation or sentiment analysis system.
- Integrate an AI tool with a custom prompt system using LangChain.
- Evaluate the ethical implications of prompt engineering (bias, cultural variation).

Module 3: Retrieval-Augmented Generation (RAG) and Multimodal AI (15 Hours)

Theory:

Concept of RAG and architecture overview. Embeddings, indexing, and vector databases (FAISS, Pinecone, ChromaDB). Generative AI tools for text, code, image, audio, and video generation. Multimodal Generative AI: Text-to-image, image-to-text, and cross-domain applications.

Practice:

- Implement a RAG-based chatbot using document embeddings.
- Create a series of prompts for a storytelling AI with contextual retrieval.
- Develop a text-to-image or text-to-audio generative demo.
- Design a user study to test the clarity and reliability of AI-generated responses.
- Analyse bias and fairness in AI-generated multimodal outputs.

Module 4: Agentic AI (15 Hours)

Theory:

Agent Architecture: Planning, tool invocation, self-reflection, and reasoning. LangChain Agents: Design, debugging, and extension methods. Overview of CrewAI, n8n, and AutoGen for multi-agent automation. Inter-agent collaboration and feedback loops. Building self-improving, autonomous AI systems.

Practice:

- Build and test a LangChain Agent or CrewAI workflow.
- Develop a prompt for a problem-solving AI agent.
- Create a prompt for a self-learning AI application.
- Multi-agent task (e.g., one agent retrieves data, another summarises).
- Evaluate performance and coordination between multiple agents.
- Evaluate ethical implications of AI-generated content.
- Analyze bias and cultural sensitivity in AI responses.
- Develop a prompt for an environmental sustainability application.
- Propose a capstone project idea for a responsible AI application.
- Create a framework for evaluating trust and transparency in AI systems.

Textbook

1. "Prompt Engineering for Generative AI" by James Phoenix, Mike Taylor O'Reilly Media, Inc.
2. Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning. MIT Press, 2016.

Reference Books:

1. Gilbert Mizrahi, "Unlocking the Secrets of Prompt Engineering: Master the Art of Creative Language Generation to Accelerate Your Journey from Novice to Pro", January 2024. https://www.packtpub.com/en-in/product/unlocking-the-secrets-of- prompt-engineering-9781835083833.
2. Michael Ferguson, "Prompt Engineering: The Future of Language Generation", January 2023. https://books.apple.com/us/book/prompt-engineering-the-future-of- language-generation/id6445529200 .
3. "Prompt Engineering Guide", https://www.promptingguide.ai/
4. "Prompt Engineering for Generative AI", Google, https://developers.google.com/machine-learning/resources/prompt-eng.
5. "Prompt Engineering", OpenAI, https://platform.openai.com/docs/guides/prompt-engineering/strategy-test-changes-systematically.

Design and Analysis of Algorithms (60 Hours)

Course	Code	T-P-Pr	Credit
Design and Analysis of Algorithms	MTDS1201	2-2-0	4

Course Objectives:

- To introduce fundamental concepts of algorithm design and analysis using formal methods and asymptotic notations.
- To develop the ability to design efficient algorithms using Divide & Conquer, Greedy, and Dynamic Programming techniques.
- To impart knowledge of advanced algorithms for graph processing, string matching, and optimization.

Course Outcomes:

At the end of the course, students will be able to:

- **CO1:** Analyze algorithms using worst-case, average-case, and asymptotic performance measures.
- **CO2:** Apply Divide & Conquer, Dynamic Programming, and Greedy strategies to solve computational problems.
- **CO3:** Implement graph algorithms including shortest path algorithms and understand their time complexities.
- **CO4:** Apply pattern matching algorithms and analyze their performance.
- **CO5:** Understand NP-completeness, perform reductions, and apply approximation algorithms to NP-hard problems.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3		2										3	2	
CO2	2	3	3										3	3	
CO3	2	3		3									3	3	2
CO4		2	3										2	3	
CO5					3				2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

The role of Algorithms in Computing, Insertion Sort, Analyzing algorithms, Designing algorithms, Asymptotic notations

Module II (15 Hours):

Divide and Conquer Technique, heapsort, merge sort, quick sort and their time complexity, Red-Black Trees
Dynamic Programming: (Matrix-chain multiplication, Longest common subsequences), Greedy Technique: An activity selection problem , Elements of greedy strategy, Huffman codes

Module III (15 Hours):

Single –Source Shortest Paths: The Bellman-Ford algorithm, Single-source shortest paths in directed acyclic graphs, Dijkstra's algorithm.
String Matching: The naïve string matching algorithm, The Rabin Karp algorithm

Module IV (15 Hours):

NP completeness, Reductions, coping with NP completeness, Approximation algorithms: The vertex cover problem, the travelling salesman problem, The set covering problem, The Subset-sum problem. Graph colouring.

Text Books:

T. H. Cormen, C. E. Leiserson, R. L. Rivest, Clifford Stein. "Introduction to Algorithms," Third edition ,Prentice Hall India, 2011

Reference Books:

1. Sara. Basse, Allen Van Gelder, "Computer Algorithms: Introduction to Design and Analysis", Pearson, 2000.
2. R. Motwani and P. Raghavan, "Randomized Algorithms," Cambridge University Press, 1995.
3. Dexter C .Kozen, "The Design and Analysis of Algorithms," Springer, 1992.

Big Data Systems (60 Hours)

Course	Code	T-P-Pr	Credit
Big Data Systems	MTDS1202	2-2-0	4

Course Objectives:

- To introduce the characteristics, challenges, and analytical requirements of Big Data systems.
- To develop understanding of Hadoop architecture, components, and MapReduce programming.
- To familiarize students with the Hadoop ecosystem including serialization, coordination, databases, scripting, and streaming tools.

Course Outcomes:

At the end of the course, students will be able to:

- **CO1:** Understand the characteristics of Big Data, analytical architectures, and challenges in Big Data processing.
- **CO2:** Apply Hadoop framework concepts, HDFS operations, and MapReduce programming techniques.
- **CO3:** Utilize Hadoop ecosystem tools such as AVRO, Zookeeper, HBase, Hive, Pig, Flink, and Storm for Big Data workflows.
- **CO4:** Apply GPU computing concepts and implement basic parallel programs using the CUDA environment.
- **CO5:** Analyze and evaluate Big Data frameworks for large-scale data storage, processing, and computational efficiency.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2											3	2	
CO2	2	3	3										3	3	
CO3	2	3		2									3	3	2
CO4			3		3								2	3	
CO5		2							2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

Data Storage and Analysis - Characteristics of Big Data – Big Data Analytics - Typical Analytical Architecture – Requirement for new analytical architecture – Challenges in Big Data Analytics – Need of big data frameworks

Module II (15 Hours):

Hadoop – Requirement of Hadoop Framework - Design principle of Hadoop –Comparison with other system - Hadoop Components – Hadoop 1 vs Hadoop 2 – Hadoop Daemon's – HDFS Commands – Map Reduce Programming: I/O formats, Map side join, Reduce Side Join, Secondary sorting, Pipelining MapReduce jobs

Module III (15 Hours):

Introduction to Hadoop ecosystem technologies: Serialization: AVRO, Co-ordination: Zookeeper, Databases: HBase, Hive, Scripting language: Pig, Streaming: Flink, Storm

Module IV (15 Hours):

Introduction to GPU Computing, CUDA Programming Model, CUDA API, Simple Matrix, Multiplication in CUDA, CUDA Memory Model, Shared Memory Matrix Multiplication, Additional CUDA API Features.

Text Books:

Mohammed Guller, Big Data Analytics with Spark, Apress,2015 Reference Books:

Reference Books:

1. Mike Frampton, "Mastering Apache Spark", Packt Publishing, 2015.
2. TomWhite,"Hadoop:TheDefinitiveGuide",O'Reilly,4thEdition,2015.
3. Nick Pentreath, Machine Learning with Spark, Packt Publishing, 2015.
4. Donald Miner, Adam Shook, "Map Reduce Design Pattern", O'Reilly, 2012

Digital Image Processing (60 Hours)

Course	Code	T-P-Pr	Credit
Digital Image Processing	MTDS1203	0-4-0	4

Course Objectives:

- To introduce the fundamental concepts and components of a digital image processing system.
- To develop understanding of spatial and frequency domain techniques for image enhancement and restoration.
- To provide in-depth knowledge of image segmentation techniques including edge detection, thresholding, region-based methods, and morphological approaches.

Course Outcomes:

At the end of the course, students will be able to:

- **CO1:** Understand the basic concepts, operations, and components of digital image processing systems.
- **CO2:** Apply spatial and frequency domain filtering techniques for image enhancement and restoration.
- **CO3:** Analyze and process color images using different color models and transformations.
- **CO4:** Implement various image segmentation techniques including edge detection, thresholding, and region-based methods.
- **CO5:** Apply morphological and motion-based segmentation approaches for complex image analysis tasks.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3		2										3	2	
CO2	2	3	3										3	3	
CO3	2	3		2									3	3	2
CO4		2	3		3								2	3	
CO5									2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

Basic concepts of digital image processing, Steps in Digital image processing, components of an image processing system. Histogram processing, spatial filtering, smoothing spatial filters, sharpening spatial filters.

Module II (15 Hours):

Noise models, Restoration in the presence of noise, periodic noise reduction by frequency domain filtering, linear, position-invariant degradations, estimating the degradation function. Inverse filtering, minimum mean square error filtering, constrained least square filtering, geometric mean filter, image reconstruction from projections.

Module III (15 Hours):

Color models, Pseudo color image processing, basics of full color image processing, color transformations, smoothing and sharpening, image segmentation based on color, color image compression.

Module IV (15 Hours):

Image segmentation, point, line and edge detection, thresholding, region based segmentation, segmentation using morphological watersheds, the use of motion in segmentation

Text Books:

Digital Image Processing, Rafeal C.Gonzalez, Richard E.Woods, Second Edition, Pearson Education/PHI.

Reference Books:

1. Image Processing, Analysis, and Machine Vision, Milan Sonka, Vaclav Hlavac and Roger Boyle, Second Edition, Thomson Learning.
2. Introduction to Digital Image Processing with Matlab, Alasdair McAndrew, Thomson Course Technology
3. Computer Vision and Image Processing, Adrian Low, Second Edition, B.S.Publications.
4. Digital Image Processing using Matlab, Rafeal C.Gonzalez, Richard E.Woods, Steven L. Eddins, Pearson Education.

Information Retrieval (60 Hours)

Course	Code	T-P-Pr	Credit
Information Retrieval	MTDS1204	2-2-0	4

Course Objectives:

- To introduce fundamental concepts of information retrieval, information needs, relevance, and retrieval models.
- To develop understanding of text preprocessing techniques including tokenization, stemming, stop-word removal, and dictionary construction.
- To equip students with knowledge of retrieval evaluation methods, vector space models, and similarity measures.

Course Outcomes:

At the end of the course, students will be able to:

- **CO1:** Understand the foundations of information retrieval, IR models, indexing, and ranked retrieval techniques.
- **CO2:** Apply text preprocessing, dictionary construction, and tolerant retrieval methods for efficient search and retrieval.
- **CO3:** Evaluate retrieval systems using precision, recall, F-measure, and other standard metrics.
- **CO4:** Implement document representation, vector space models, and similarity-based ranking.
- **CO5:** Apply supervised and unsupervised methods including Naive Bayes, KNN, K-Means, and basic recommendation techniques in IR applications.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3		2										3	2	
CO2	2	3	3										3	3	
CO3	2	3		2									3	3	2
CO4		2	3		3								2	3	
CO5									2				2	2	3

***High-3, Medium-2, Low-1**

Module I (15 Hours):

Introduction to Retrieval and IR Model: Information, Information Need and Relevance; The IR System; Boolean Retrieval; Term Vocabulary and Postings list; Index Construction; Ranked and other alternative Retrieval Models.

Module II (15 Hours):

Dictionary and Tolerant Retrieval: Tokenization, Stop words, Stemming, Inverted index, Wild card queries, Jaccard coefficient, Understand the importance of Scoring, term weighting.

Module III (15 Hours):

Retrieval Evaluation: Precision, Recall, F-measure, E-measure, Normalized recall, Evaluation problems, Document Processing: Representation; Vector Space Model; Feature Selection; Stop Words; Stemming; Notion of Document Similarity; Standard Datasets.

Module IV (15 Hours):

Classification and Clustering: Notion of Supervised and Unsupervised Algorithms; Naive Bayes, Nearest Neighbour; Clustering Methods such as K-Means, Introduction to recommendation system Collaborative , Content based recommendation.

Text Books:

1. Introduction to Information Retrieval , Christopher D. Manning and Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press, 2008.

Reference Books:

1. Information Storage and Retrieval Systems: Theory and Implementation, Kowalski, Gerald, Mark T Maybury, Springer.
2. Modern Information Retrieval, Ricardo Baeza-Yates, Pearson Education, 2007.
3. Information Retrieval: Algorithms and Heuristics, David A Grossman and Ophir Frieder, 2nd Edition, Springer, 2004.
4. Information Retrieval Data Structures and Algorithms, William B Frakes, Ricardo BaezaYates, Pearson Education, 1992.

Computational Intelligence (60 Hours)

Course	Code	T-P-Pr	Credit
Computational Intelligence	MTDS1205	2-2-0	4

Course Objectives:

- To introduce the concepts, architectures, and learning methods of Artificial Neural Networks (ANN).
- To develop understanding of fuzzy set theory, fuzzy logic systems, and fuzzy rule-based reasoning.
- To impart foundational knowledge of Genetic Algorithms (GA), their operators, modeling, and applications.

Course Outcomes:

At the end of the course, students will be able to:

- **CO1:** Understand the fundamentals, models, and learning methods of Artificial Neural Networks.
- **CO2:** Apply fuzzy set theory, fuzzy logic, and fuzzy rule-based reasoning to design inference systems.
- **CO3:** Implement genetic algorithms, including encoding, fitness evaluation, and genetic operators for optimization.
- **CO4:** Analyze and develop hybrid soft computing systems combining ANN, fuzzy logic, and GA.
- **CO5:** Apply computational intelligence techniques to real-life engineering and optimization problems.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3		2										3	2	
CO2	2	3	3										3	3	
CO3	2	3		2									3	3	2
CO4		2	3		3								2	3	
CO5									2				2	2	3

***High-3, Medium-2, Low-1**

Module - I (15 Hrs.)

Artificial Neural Network(ANN): Fundamentals of ANN, Basic Models of an artificial Neuron, Neural Network Architecture, Learning methods, Terminologies of ANN, Supervised Learning Networks: Perceptron, MLP, Architecture of a Back propagation Network : back propagation

Module –II (15 Hrs)

Fuzzy set theory: crisp sets, fuzzy sets, crisp relations, fuzzy relations, Fuzzy Systems: Crisp logic predicate logic, fuzzy logic, fuzzy Rule based system, Defuzzification Methods, Fuzzy rule based reasoning

Module –III (15 Hrs)

Fundamentals of genetic algorithms: Encoding, Fitness functions, Reproduction. Genetic Modeling: Cross cover, Inversion and deletion, Mutation operator, Bit-wise operators, Bitwise operators used in GA. Convergence of Genetic algorithm. Applications, Real life Problems.

Module – IV (15 Hrs.)

Hybrid Soft Computing Techniques Hybrid system, neural Networks, fuzzy logic and Genetic algorithms hybrids. Genetic Algorithm based Back propagation Networks: GA based weight determination applications: Fuzzy logic controlled genetic Algorithms soft computing tools, Applications.

Text Book:

Principles of Soft Computing- S.N.Sivanandan and S.N.Deepa, Wiley India, 2ndEdition,2011

Reference Book:

1. Neuro Fuzzy and Soft Computing, J. S. R. JANG,C.T. Sun, E. Mitzutani, PHI
2. Neural Networks, Fuzzy Logic, and Genetic Algorithm (synthesis and Application) S.Rajasekaran, G.A. VijayalakshmiPai, PHI

Research Methodology and IPR (75 Hours)

Course	Code	T-P-Pr	Credit
Research Methodology and IPR	CUTM3191	0-2-2	4

Course Objectives:

- To develop an understanding of various research designs and techniques.
- To identify various sources of information for literature review and data collection.
- To develop an understanding of the ethical dimensions of conducting applied research.

Course Outcomes:

At the end of the course, students will be able to:

- **CO1** Understand the principles of research design and data collection methods as well as the legal framework of intellectual property rights**.**
- **CO2** Develop research proposals and formulate research questions**.**
- **CO3** Identify different types of intellectual property such as patents, trademarks, copyrights, and trade secrets.
- **CO4** Develop the ability to critically evaluate research studies with appropriate techniques, resources which will lead to evaluate the strengths and limitations of different types of intellectual property protection.
- **CO5** Develop the ability to critically evaluate research studies with appropriate techniques, resources which will lead to evaluate the strengths and limitations of different types of intellectual property protection.

Course Outcome to Program Outcome Mapping:

COs / POs / PSOs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3		3												
CO2				3		2									
CO3		2		3											
CO4					2										
CO5							2	3							

***High-3, Medium-2, Low-1**

Module I

Elementary Research Methodology (15 Hours)

Research Concept, Objective, characteristics, Steps and Significance of Research, Arbitrary and Scientific Research, Research approaches. Types of research: Historical, Descriptive, Analytical, Case Study, Quantitative vs. qualitative, Conceptual, Empirical Action Research, Research Methods vs Methodology. Research Problems: Selection and definition of the research problems, formulating a research problem, identifying variables and Constructing hypothesis; Choosing a mentor, lab and research question; maintaining a lab notebook; Selection of problems - stages in the execution of research.

Module II

Academic Writing and Presentation (15 Hours)

Technical writing skills - types of reports; layout of a formal report; standard of Journal (Impact Factor, Citation Index), Scientific writing skills - importance of communicating science; problems while writing a scientific document; plagiarism, software for plagiarism; scientific publication writing: elements of a scientific paper including abstract, introduction, materials & methods, results, discussion, references; drafting titles and framing abstracts; publishing scientific papers - peer review process and problems, recent developments such as open access and non-blind review; characteristics of effective technical communication; scientific presentations; ethical issues; scientific misconduct.

Module III

Scientific Communication Skills (15 Hours)
Concept of effective communication- setting clear goals for communication; determining outcomes and results; barriers to effective communication; non-verbal communication- importance of body language, power of effective listening; Presentation skills - formal presentation skills; preparing and presenting using over-head projector, PowerPoint; defending interrogation; scientific poster preparation & presentation; participating in group discussions; Computing skills for scientific research - web browsing for information search.

Module IV

Introduction to IPR (15 Hours)
Introduction to intellectual property; types of IP: patents, trademarks, copyright & related rights, industrial design, traditional knowledge, geographical indications, protection of new GMOs; IP as a factor in R& D; IPs of relevance to biotechnology and few case studies; plant variety protection and farmers rights.

Module V

Types of Patents (15 Hours)
Basics of patents: types of patents; Indian Patent Act 1970; recent amendments; WIPO Treaties; Budapest Treaty; Patent Cooperation Treaty (PCT) and implications; filing of a patent application; role of a Country Patent Office; precautions before patenting-disclosure/non-disclosure – patent application- forms and guidelines including those of National Bio-diversity

Authority (NBA) and other regulatory bodies, fee structure, time frames; types of patent applications: provisional and complete specifications.

Projects

1. Write a review article and submit to a journal
2. Write a book chapter/ book for publishing
3. Write an original article for a journal

Text Books”

1. Geoffrey Marczyk, David DeMatteo, David Festinger (2005) Essentials of Research Design and Methodology, John Wiley & Sons, Inc.
2. Carol Ellison (2010) McGraw-Hill’s ConcisevGuide to WritingvResearch Papers, McGraw-Hill
3. Kothari CR (2016) Research Methodology: Methods and Techniques, New Age Pvt Ltd
4. Ganbawale RM, (2017) Bioststistics and Research Methodology, New Central Book Agency

Reference Books:

1. Sinha, S.C. and Dhiman, A.K., (2002). Research Methodology, Ess Ess Publications. 2 volumes.
2. Trochim, W.M.K., (2005). Research Methods: the concise knowledge base, Atomic Dog
3. Publishing. 270p.
4. Wadehra, B.L. (2000). Law relating to patents, trademarks, copyright designs and geographical indications. Universal Law Publishing.
5. Neuman, W.L. (2008). Social research methods: Qualitative and quantitative approaches, Pearson Education