Semester: 3	
Programme: MSc Data Science	
Course: Optimization and Data Governance and Compliance	
Paper code: M2C4DS26031CT / M2C4DS26031CP	**Credits: 6**
Hours/week: 6 (4Th + 2Pr)	
Category: Core	
Theory / Practical / Composite: Composite	
No of Module: 2	

Course Outcome:

1. **Remember** the fundamental definitions, mathematical formulations, and principles of optimization theory (convexity, search methods, KKT conditions) alongside the core frameworks of data ecosystems, data quality dimensions, and regulatory compliance.
2. **Understand** the conceptual logic underlying numerical optimization algorithms (gradient descent variants, Newton's methods, penalty-based regularization) as well as modern data architecture paradigms (Data Lakehouses, Data Mesh), privacy-enhancing technologies, and data stewardship.
3. **Apply** line search algorithms, penalty-based techniques (Ridge, LASSO, Elastic-net), and data validation or privacy-preserving pipelines (anonymization, differential privacy) to model data workflows and solve optimization problems.
4. **Analyze** constrained and unconstrained problems using Karush-Kuhn-Tucker (KKT) theory while examining modern data ecosystems to identify model performance bottlenecks, data quality anomalies, and regulatory compliance gaps.
5. **Evaluate** the convergence and efficiency of numerical optimization techniques (such as stochastic gradient descent and simulated annealing) and judge the effectiveness of data governance, security audits, and regulatory frameworks (DPDP Act, GDPR, EU AI Act).
6. **Create** integrated, production-ready data science frameworks by synthesizing advanced optimization models (such as Support Vector Machines) with robust, ethical data governance architectures, privacy safeguards, and AI compliance pipelines.

Prerequisites: ***Basic knowledge about any prior course***

SYLLABUS

Module/Unit	**CONTENT**	**HOURS or NUMBER OF CLASSES**	**CO Mapping**	**COGNITIVE LEVEL**
Module 1 / Unit 1	**Optimization Theory:** Constrained and Unconstrained Optimization, Optimization with Equality Constraints and Inequality Constraints, Integer programming, non-linear programming	6	CO1, CO4	K1, K4
Module 1 / Unit 2	**Convexity**: Convex Sets, Convex Functions, Minimization of Convex Functions, Convex programming	2	CO1, CO2	K1, K2
Module 1 / Unit 3	**Some numerical optimization methods**: Line search method, Newton's Methods, Conjugate Gradient and Quasi-Newton Methods, Gradient descent, Stochastic gradient descent, Mini-batch gradient descent, Simulated annealing.	11	CO1, CO3	K1, K3

Module 1 / Unit 4	**Karush-Kuhn-Tucker Theory**: The Multiplier Rule, Constraint Qualification, A Sufficient Condition for a Minimum. Primal-Dual methods	7	CO1, CO4, CO6	K1, K4, K6
Module 2 / Unit 1	**Modern Data Ecosystem & Architecture** *Updates / Enhancements*: Expanded beyond traditional databases and data lakes to include **Data Lakehouses** and **Data Mesh** paradigms, emphasizing distributed ownership, data-as-a-product, and enterprise metadata management.	4	CO1, CO2	K1, K2
Module 2 / Unit 2	**Data Quality & Integrity** *Updates / Enhancements*: Enhanced focus on automated data quality dimensions, statistical relevance, automated anomaly detection, and data validation pipelines essential for modern machine learning pipelines.	5	CO1, CO3, CO4	K1, K3, K4
Module 2 / Unit 3	**Data Governance, Ethics & Regulatory Compliance** *Updates / Enhancements*: Integrated modern regulatory frameworks (such as India's **DPDP Act**, **GDPR**, and the **EU AI Act**), alongside the FAT Flow framework (Fairness, Accountability, Transparency), consent management, and data stewardship.	6	CO1, CO2, CO5	K1, K2, K5
Module 2 / Unit 4	**Data Privacy & Confidentiality** *Updates / Enhancements*: Added advanced Privacy-Enhancing Technologies (PETs) including **Differential Privacy** and **Synthetic Data Generation**, alongside PII handling, anonymization, and balancing privacy preservation with analytical utility, ISO42001.	6	CO3, CO6	K3, K6
Module 2 / Unit 5	**Data Security, Audits & AI Compliance** *Updates / Enhancements*: Included enterprise security solutions, Data Security Audits, and emerging compliance challenges specific to **Generative AI and Large Language Models (LLMs)** (e.g., copyright, data provenance, and model guardrails).	5	CO5, CO6	K5, K6

Text Books

1. Kenneth Lange; Optimisation, Springer; 2004
2. Kenneth Lange; Numerical Analysis for statisticians, Springer, Second Edition.
3. Wenyu sun, ya-xiang yuan; Optimization Theory and Methods Nonlinear Programming; Springer, 2006.
4. Data Ethics and Challenges: Shukla S. etal, Springer Series in Applied Sciences and Technologies, 2022.
5. Ethics of Data and Analytics- Concepts and Cases: Kristen Martin, CRC Press, 2022.

6. Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools; Rupa Mahanti.		
Online Resources for Optimization course		
1. NPTEL Lectures: https://nptel.ac.in/courses/106106245		
2. https://github.com/ebrahimpichka/awesome-optimization#convex-optimization		
Evaluation	CIA: Class Performance: 2 Attendance: 3 Mid-term Exam: 10 Total: 15 End Semester Exam: Module I= 23 Module II=22 Total: 45	Continuous Assessment: 40
Paper Structure for Theory Semester Exam Module:	Short Questions	Long Questions: 10 marks each
Module I	1 out of 2 Questions, 3 marks each	2 out of 3
Module II	1 out of 2 Questions, 2 marks each	2 out of 3

Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
CO1	**Remember** the fundamental definitions, mathematical formulations, and principles of optimization theory (convexity, search methods, KKT conditions) alongside the core frameworks of data ecosystems, data quality dimensions, and regulatory compliance.	K1
CO2	**Understand** the conceptual logic underlying numerical optimization algorithms (gradient descent variants, Newton's methods, penalty-based regularization) as well as modern data architecture paradigms (Data Lakehouses, Data Mesh), privacy-enhancing technologies, and data stewardship.	K2
CO3	**Apply** line search algorithms, penalty-based techniques (Ridge, LASSO, Elastic-net), and data validation or privacy-preserving pipelines (anonymization, differential privacy) to model data workflows and solve optimization problems.	K3
CO4	**Analyze** constrained and unconstrained problems using Karush-Kuhn-Tucker (KKT) theory while examining modern data ecosystems to identify model performance bottlenecks, data quality anomalies, and regulatory compliance gaps.	K4
CO5	**Evaluate** the convergence and efficiency of numerical optimization techniques (such as stochastic gradient descent and simulated annealing) and judge the effectiveness of data governance, security audits, and regulatory frameworks (DPDP Act, GDPR, EU AI Act).	K5
CO6	**Create** integrated, production-ready data science frameworks by synthesizing advanced optimization models (such as Support Vector Machines) with robust, ethical data governance architectures, privacy safeguards, and AI compliance pipelines.	K6