Department of Computer Science and Engineering
UNIVERSITY OF KALYANI, KALYANI 741235
WEST BENGAL

SYLLABUS

FOR

MASTER OF SCIENCE (M.Sc.) in DATA SCIENCE
(Two-Year Programme)

(Effective from the Session: 2021-22)

Paper Code Convention: MDS-W–XYZ
[W=COR/AECC/GEC/DSE/SEC/DISS/GV; X = 1/2/3/4 (semester), Y = 0/1/2/3 (theory/practical/Dissertation/viva), Z (paper id)]

Paper Code	Paper Name	Type	Credit	Weekly hours (L+T+P)	Marks Theory - (Exam+Internal Assessment)
Semester - I					
MDS-COR–101	Mathematical Foundation	Theory	4	3+1+0	100 (80+20)
MDS-COR–102	Statistics and Data Analytics	Theory	4	3+1+0	100 (80+20)
MDS-COR–103	Algorithms and Data Structure	Theory	4	3+1+0	100 (80+20)
MDS-COR–104	Database Management Systems	Theory	4	3+1+0	100 (80+20)
MDS-COR–105	Operating Systems	Theory	4	3+1+0	100 (80+20)
MDS-COR–111	Statistics and Data Analytics Laboratory	Practical	2	0+0+2	50
MDS-COR–112	Algorithms & Data Structure Laboratory	Practical	3	0+0+3	100
MDS-COR–113	Database Management Systems Laboratory	Practical	3	0+0+3	100
MDS-AECC–114	Communicative English and Business Presentation	Practical	2	0+0+2	50
		TOTAL	30	30	800
Semester - II					
MDS–GEC-201	CBCS Open Choice Course	Theory	4	3+1+0	100 (80+20)
MDS–COR-202	Scientific Computing	Theory	4	3+1+0	100 (80+20)
MDS–COR-203	Data Warehousing and Data Mining	Theory	4	3+1+0	100 (80+20)
MDS–COR-204	Artificial Intelligence	Theory	4	3+1+0	100 (80+20)
MDS–DSE-205	Elective – I	Theory	4	3+1+0	100 (80+20)
MDS–COR-211	Scientific Computing Laboratory	Practical	3	0+0+3	100
MDS–COR-212	Data Mining and Visualization Laboratory	Practical	3	0+0+3	100
MDS–COR-213	Artificial Intelligence Laboratory	Practical	3	0+0+3	100
		TOTAL	29	29	800
Semester - III					
MDS-COR–301	Machine Learning and Deep Learning	Theory	4	3+1+0	100 (80+20)
MDS-SEC–302	Big Data Algorithms & Analytics	Theory	4	3+1+0	100 (80+20)
MDS-DSE–303	Elective – II	Theory	4	3+1+0	100 (80+20)
MDS-DSE-304	Elective – III	Theory	4	3+1+0	100 (80+20)
MDS-COR–311	Machine Learning and Deep Learning Laboratory	Practical	3	0+0+3	100
MDS-SEC–312	Big Data Algorithms & Analytics Laboratory	Practical	3	0+0+3	100
MDS-DSE-313	Advanced Programming Laboratory	Practical	3	0+0+3	100
MDS-DISS–321	Dissertation-I	Dissertation	6	0+0+6	100
		TOTAL	31	31	800
Semester - IV					
MDS-DISS–421	Dissertation-II	Dissertation	12	0+0+20	300
MDS-GV–431	Grand Viva	Viva	4	-	100
		TOTAL	16	20	400
		Overall Total	106	110	2800

COR: Core Courses
AECC: Ability Enhancement Compulsory Courses
GEC: Generic Elective Courses
SEC: Skill Enhancement Courses
DSE: Discipline Specific Elective
DISS: Dissertation
GV: Grand Viva

Elective Papers

1. Time Series Analysis
2. Multi-criteria Decision Making
3. Speech & Natural Language Processing
4. Information Security
5. Computational Biology
6. Mobile Applications & Programming
7. Cloud Computing
8. Digital Image Processing and Computer Vision
9. Web-based Programming
10. Soft Computing
11. Complex Network Analysis
12. Operations Research
13. High Performance Computing
14. Blockchain Technology

* New elective papers may included/offered as per the need of the industry and modern technologies as and when required with the approval of the PG-BoS of Dept. of Computer Science & Engineering.

Semester I

MDS-COR–101. Mathematical Foundation
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Functions and Relations: Definition and type of functions, mappings, injection, bijection and surjection, equivalence relations and partitions, partial ordering relation, Lattices and their applications, generating functions, recurrence relations, solution of linear homogeneous and non- homogeneous recurrence relations by the method of generating functions and particular solution method.
Linear Algebra: Basic properties of matrices and vectors: scalar multiplication, linear transformation, transpose, conjugate, rank, determinant, Inner and outer products, matrix multiplication, matrix inverse, unit vectors, symmetric matrix, matrix factorization concept/LU decomposition, Gaussian/Gauss-Jordan elimination, solving Ax=b linear system of equation, Vector space, basis, span, orthogonality, orthonormality, linear least square, eigenvalues, eigenvectors.
Algebraic Structures: Groups, Rings and Fields, Group Codes, Concepts of Vector Spaces.
Calculus: Functions of a single variable, limit, continuity, differentiability, Mean value theorems, indeterminate forms, L'Hospital's rule, Maxima and minima, Product and chain rule, Taylor's series, infinite series summation/integration concepts, Fundamental and mean value-theorems of integral calculus, evaluation of definite and improper integrals, Beta and gamma functions, Functions of multiple variables, limit, continuity, partial derivatives, Basics of ordinary and partial differential equations.
Discrete Mathematics: Sets, subsets, power sets, set operations, Counting functions, combinatorics, countability, basic proof techniques: induction, proof by contradiction, Basics of inductive, deductive, and propositional logic, Graphs theory.

1. Narsingh Deo, Graph Theory With Applications To Engineering And Computer Science, PHI Learning
2. C. L. Liu, Elements of Discrete Mathematics, TMH, 2000.
3. Kenneth H. Rosen; Discrete Mathematics and its applications; TMH.
4. K. H. Rosen, Discrete Mathematics and applications, fifth edition 2003, TMH.
5. Ross, S., A First Course in Probability, Collier Macmillan, New York, 1976
6. Liu, C.L., Introduction to Combinatorial Mathematics, McGraw Hill. 1996
7. R.P. Grimaldi, B. V. Ramana, Discrete and Combinatorial mathematics: An applied introduction, Pearson Education, 2007
8. Murray, R., J. Spiegel, and R. Schiller. Schaum's outline of probability and statistics. 2013.
9. Lipschutz, Seymour, and Marc Lars Lipson. Discrete mathematics. McGraw-Hili, 2007.

MDS-COR–102. Statistics and Data Analytics
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction to data analytics, type of data, data collection, blending and preparation, data preprocessing, ETL techniques, data mining, data science project life cycle; Data summaries and descriptive statistics, measures of central tendency, variance, covariance, correlation; Basic probability: basic idea, expectation, probability calculus, Bayes' theorem, conditional probability; Probability distributions: uniform, normal, exponential, binomial, chi-square, Student's t-distribution, F-distribution, central limit theorem; Sampling, measurement, error, random sample generation; Hypothesis testing, A/B testing, confidence intervals, p-values; ANOVA, Post hoc analysis, t- test, Z-

test, F-test; Non-parametric & distribution free statistics - Sign test, Wilcoxon signed-rank test, Spearman's rank-order correlation, Multifactor dimensionality reduction (MDR), Principal Component Analysis (PCA); Regression, bias variance, Linear Regression, Multivariate Regression, Subset Selection, Shrinkage Methods, Principal Component Regression, Partial Least squares, Logistic Regression, LDA; Gaussian Mixture Models, Expectation Maximization; Introduction to Time Series Analysis, ARMA models.

Text Books:

1. Cady, Field. The data science handbook. John Wiley & Sons, 2017.
2. Peter Bruce and Andrew Bruce, Practical Statistics for Data Science, O'Reilly, 2017.
3. Moreira, João, Andre Carvalho, and Tomás Horvath. A general introduction to data analytics. John Wiley & Sons, 2018.
4. J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann/Elsevier India, 2001.
5. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. No. 10. New York: Springer series in statistics, 2001.

Reference Books:

1. Alpaydin, Ethem. Introduction to machine learning. MIT press, 2009.
3.D. Hand, H. Mannila, and P. Smyth. Principles of Data Mining, MIT Press, 2001.
4.Pujari, Arun K. Data mining techniques. Universities press, 2001.

MDS-COR–103. Algorithms and Data Structure
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Data Types and Algorithms: Time and Space Analysis Of Algorithms-Order Natation's: Linear Data Structures: Sequential Storage Representation-Arrays, Strings, Stacks, Queues, Dequeues and other their Applications: Linear Data Structures: Linked Storage Lists, Circularly Linked Lists, Doubly Linked Lists, Applications: Recursion-Design of Recursive Algorithms, Tail Recursion, When Not to use Recursion, Removal of Recursion; Non-Linear Data Structures: Trees, Binary Trees, Binary Search Tree, Traversals and Threads, Insertion and Deletion Algorithms, Height-Balanced and Weight-Balanced Trees, B-Trees, B+ Trees, Applications of Trees: Graphs-Representation, Sorting and Searching-Review of Various Algorithms, Hashing.

Text Books:

1. Ellis Horowitz, S. Sahni, D. Mehta Fundamentals of Data Structures in C++, Galgotia Book Source, New Delhi.
2. T.H. Cormen, C.E. Leiserson, R.L. Rivest and C. Stein, Introduction to Algorithms (2nd ed.), Prentice-Hall of India, 2006
3. Y. Langsam, M. Augenstin and A. Tannenbaum, Data Structures using C and C++, Pearson Education Asia, 2nd Edition, 2002.
4. Aho Alfred V., Hopperoft John E., UIlman Jeffrey D., "Data Structures and Algorithms", Addison Wesley

Reference Books:

1. Debasis Samanta, Classic Data Structures, PHI, 2nd Edition
2. S. Lipschutz, Data Structures Mc-Graw Hill International Editions, 1986.
3. Jean-Paul Tremblay, Paul. G. Soresan, An introduction to data structures with Applications, Tata Mc-Graw Hill International Editions, 2nd edition, 1984.
4. A. Michael Berman, Data structures via C++, Oxford University Press, 2002.
5. M. Weiss, Data Structures and Algorithm Analysis in C++, Pearson Education, 2002, 2nd edition.
6. M.T. Goodrich, R. Tamassia and D. Mount, Data Structures and Algorithms in C++, John Wiley & Sons, Inc., 2004.
7. M.J. Folk, B. Zoellick and G. Riccardi, File Structures: An Object Oriented Approach With C++ (3rd ed.), Addison- Wesley, 1997.
8. Robert L. Kruse and A.J. Ryba, Data structures and program design in C++, Prentice-Hall, Inc., NJ, 1998.
9. B. Stroupstrup, The C++ Programming Language, Addison Wesley, 2004.
10. D.E.Knuth, Fundamental Algorithms, Vol. I, Addison Wesley, 1997.

MDS-COR–104. Database Management Systems
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction: Concept & Overview of DBMS, Data Models, Database Languages, Database Administrator, Database Users, Three Schema architecture of DBMS.
Entity-Relationship Model: Basic concepts, Design Issues, Mapping Constraints, Keys, Entity-Relationship Diagram, Weak Entity Sets, Extended E-R features.
Relational Model: Structure of relational Databases, Relational Algebra, Relational Calculus, Extended Relational Algebra Operations,
Views, Modifications of the Database.
SQL and Integrity Constraints: Concept of DDL, DML, DCL. Basic Structure, Set operations, Aggregate Functions, Null Values, Domain Constraints, Referential Integrity Constraints, assertions, views, Nested Subqueries, Database security application development using SQL, Stored procedures and triggers.
Relational Database Design: Functional Dependency, Different anomalies in designing a Database., Normalization using functional dependencies, Decomposition, Boyce-Codd Normal Form, 3NF, Normalization using multi-valued dependencies, 4NF, 5NF.
Internals of RDBMS: Physical data structures, Query optimization: join algorithm, statistics and cost bas optimization. Transaction processing, Concurrency control and Recovery Management: transaction model properties, state serializability, lock base protocols, two phase locking.
File Organization & Index Structures: File & Record Concept, Placing file records on Disk, Fixed and Variable sized Records, Types of Single-Level Index (primary, secondary, clustering), Multilevel Indexes, Dynamic Multilevel Indexes using B tree and B+ tree.

Text Books:

1. Henry F. Korth and Silberschatz Abraham, “Database System Concepts”, Mc.Graw Hill.
2. Elmasri Ramez and Novathe Shamkant, “Fundamentals of Database Systems”, Benjamin Cummings Publishing. Company.
3. Ramakrishnan: Database Management System, McGraw-Hill
4. Gray Jim and Reuter Address, “Transaction Processing: Concepts and Techniques”, Moragan

Kauffman
Publishers.
5. Jain: Advanced Database Management System CyberTech
6. Date C. J., "Introduction to Database Management", Vol. I, II, III, Addison Wesley.
7. Ullman JD., "Principles of Database Systems", Galgottia Publication.

Reference Books:
1. James Martin, "Principles of Database Management Systems", 1985, Prentice Hall of India, New Delhi
2. "Fundamentals of Database Systems", RamezElmasri, ShamkantB.Navathe, Addison Wesley Publishing
Edition
3. "Database Management Systems", ArunK.Majumdar, Pritimay Bhattacharya, Tata McGraw Hill

MDS-COR–105. Operating Systems
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Operating System concepts - OS Structure – OS Services - System Calls. Process management: Process Concept - Operations on process - Cooperating processes- Inter-process communication. Process scheduling - Scheduling algorithms.
Threads- Multithreading models – Containers - Process synchronization- critical-section – Synchronization hardware – Semaphores – Classic problems of synchronization – critical regions.
Deadlocks: Characterization, Prevention, Avoidance, Detection, and Recovery.
Memory Management: Paging, segmentation, Demand Paging, Page Replacement, Allocation of Frames.
File Concepts, Access, and Allocation Methods, Free Space Management. Disk Structure; Disk Scheduling, and Disk Management.
Virtual Machines: Types of Virtual Machines and Implementations; Virtualization.
Linux Operating Systems: Design Principles, Kernel Modules, Process Management, Scheduling, Memory Management, File Systems, Input and Output; Inter-process Communication, Network Structure.
Windows Operating Systems: Design Principles, System Components, Terminal Services and Fast User Switching; File System, Networking.
Distributed Systems: Types of Network based Operating Systems, Network Structure, Communication Structure and Protocols; Robustness, Design Issues, Distributed File Systems.

Books:
1. Abraham Silberschatz, Peter B. Galvin and Greg Gagne, "Operating System Concepts Essentials", John Wiley & Sons Inc., 2010.
2. Andrew S. Tanenbaum, "Modern Operating Systems", 3rd Edition, Prentice Hall, 2007.
3. William Stallings, "Operating Systems: Internals and Design Principles", 7th Edition, Prentice Hall, 2011.
4. Garry Nutt, "Operating Systems", 3rd Edition, Addison-Wesley, 2003.
5. Dietel H. N., "An Introduction to Operating Systems", Addison Wesley.

MDS-COR–111. Statistics and Data Analytics Laboratory
Full Marks: 50, Weekly Hours: 0 + 0 + 2
Allotted Hrs: 20P

R/Python/Matlab programming
Statistical analysis using R/Excel/SPSS/Python/Matlab

Books Recommended

1. Gardener, M (2012) Beginning R: The Statistical Programming Language, Wiley Publications.
2. Braun W J, Murdoch D J (2007): A First Course in Statistical Programming with R. Cambridge University Press. New York.
3. Moore, D.S. and McCabe, G.P. and Craig, B.A. (2014): Introduction to the Practice of Statistics, W.H. Freeman
4. Cunningham, B.J (2012):Using SPSS: An Interactive Hands-on approach
5. Cho, M,J., Martinez, W.L. (2014) Statistics in MATLAB: A Primer, Chapman and Hall/CRC Brett Lantz, Machine Learning with R.
6. Cory Lesmeister, Mastering Machine Learning with R.

MDS-COR–112. Algorithms & Data Structure Laboratory
Full Marks: 100, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P

Lab pertaining to MDS-COR-103

MDS-COR–113. Database Management Systems Laboratory
Full Marks: 75, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P

Lab pertaining to MDS-COR-104

MDS-AECC–114. Communicative English and Business Presentation
Full Marks: 50, Weekly Hours: 0 + 0 + 2
Allotted Hrs: 20P

Semester II

MDS–GEC-201. CBCS Open Choice Course (For other departments/courses)
Introduction to Data Science
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction, What is Data Science? – The core problems and solutions. Extracting Intelligence from Data – formulating problems, the Data Pipeline, Types of Data, Data collection, Data storage, Data warehousing, Data Wrangling, ETL, Cleaning and Preparation, Data Mining, KDD, Data Presentation,

Basic concepts in Statistics and Exploratory Data Analysis, basic concepts of data structures, basic concepts of database management systems.
Introduction to machine learning, supervised and unsupervised learning, regression – linear regression, logistic regression, classification – k-nearest neighbor, decision tree, random forest, neural networks, SVM; clustering – K-means, hierarchical;association rule mining - apriori algorithm, feature extraction, dimensionality reduction, Multidimensional scaling (MDS), PCA, SVD, feature selection. Applications of Data Science and Artificial Intelligence in banking and finance, biology and healthcare, computer vision, natural language processing etc.

Books Recommended

1. Cady, Field. *The data science handbook*. John Wiley & Sons, 2017.
2. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann/Elsevier India, 2001.
3. Pierson, Lillian. *Data science for dummies*. John Wiley & Sons, 2015.

MDS-COR–202. Scientific Computing
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Definition and sources of errors, solutions of nonlinear equations; Bisection method, Newton's method and its variants, fixed point iterations, convergence analysis; Newton's method for non-linear systems; Finite differences, polynomial interpolation; Numerical integration - Trapezoidal and Simpson's rules, Gaussian quadrature; Initial value problems - Taylor series method, Euler and modified Euler methods, Runge-Kutta methods.
Linear programming, simplex algorithm, Integer programming, Constraint programming, knapsack problem

Text/Reference Books:

1. D. Kincaid and W. Cheney, Numerical Mathematics and Computing, 7th Edn., Cengage, 2013. K. E.
2. Atkinson, Introduction to Numerical Analysis, 2nd Edn., John Wiley, 1989.
3. Rajaraman, Vaidyeswaran. *Computer oriented numerical methods*. PHI Learning Pvt. Ltd., 2018.
4. F.S. Hillier and G.J. Lieberman: Introduction to Operations Research- Concepts and Cases, 9th Edition, Tata McGraw Hill, 2010.
5. An introduction to Optimization by Edwin P K Chong, Stainslaw Zak

MDS-COR–203. Data Warehousing and Data Mining
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

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 (OLAP) – OLAP and Multidimensional Data Analysis.

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.
Cluster Analysis: - Types of Data in Cluster Analysis – A Categorization of Major Clustering Methods – Partitioning Methods – Hierarchical methods – Density-Based Methods – Grid-Based Methods – Model-Based Clustering Methods – Clustering High-Dimensional Data – Constraint-Based Cluster Analysis – Outlier Analysis.
Classification and Prediction: - Issues Regarding Classification and Prediction – Nearest Neighbour Classification - Classification by Decision Tree Introduction – Bayesian Classification – Rule Based Classification –Support Vector Machines – Associative Classification –Prediction – Accuracy and Error Measures – Evaluating the Accuracy of a Classifier or Predictor – Ensemble Methods – Model Section.
Mining Object, Spatial, Multimedia, Text and Web Data:
Multidimensional Analysis and Descriptive Mining of Complex Data Objects – Spatial Data Mining – Multimedia Data Mining – Text Mining – Mining the World Wide Web.

Books:

1. Reema Thareja, “Data Warehousing”, Oxford University Press.
2. Jiawei Han and Micheline Kamber, “Data Mining Concepts & Techniques”, Elsevier Pub.
3. Margret H. Dunham “Data Mining: Introductory and Advanced topics” Pearson Education
4. Paulraj Ponniah, “Data Warehousing Fundamentals”, John Wiley & Sons, Inc.
5. Vikram Pudi, P. Radha Krishana “Data Mining”, Oxford University press.

MDS-COR–204. Artificial Intelligence
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction – What is AI – Importance of AI – objectives.
Intelligent agents, state space representation, uninformed searches – BFS, DFS, IDS, informed and heuristic searches – Branch & bound, Best first, A* search; Local searches
and optimization, local and global optima, hill climbing, gradient descent, simulated annealing, genetic algorithms, Adversarial Search: Min-Max game tree
Knowledge – Its representation, Organization – Manipulation and Acquisition.
Predicate calculus in AI – First order predicate logic & its use in knowledge representation-Resolution principle. Use of resolution in reasoning and question answering.
Uncertainly Management-Fuzzy logic, Bayesian inferencing, dempster-shafer theory of beliefs, structured representation of knowledge- - Semantic networks, frames, conceptual dependency & scripts. Expert systems-rule based system architecture non-production system architecture-knowledge acquisition methods-Explanation methods-Expert system shells, Application of AI in natural language processing, speech understanding. Computer Vision, planning, etc.

Text Books:

1. Artificial Intelligence, Ritch & Knight, TMH
2. Artificial Intelligence A Modern Approach, Stuart Russel Peter Norvig Pearson

3. Introduction to Artificial Intelligence & Expert Systems, Patterson, PHI
4. Poole, Computational Intelligence, OUP
5. Logic & Prolog Programming, Saroj Kaushik, New Age International

Reference Books:

1. Expert Systems, Giarranto, VIKAS
2. Artificial Intelligence, Russel, Pearson

MDS-DSE–205. Elective - I
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

MDS-COR–211. Scientific Computing Lab
Full Marks: 100, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P
Lab pertaining to MDS-COR-202

MDS-COR–212. Data Mining and Visualization Lab
Full Marks: 100, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P
Lab pertaining to MDS-COR-203

MDS-COR–213. Artificial Intelligence Laboratory
Full Marks: 100, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P
Lab pertaining to MDS-COR-204

Semester- III

MDS-COR–301. Machine Learning and Deep Learning
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Advanced clustering methods, variants of K-means, BIRCH, DBSCAN, Expectation-Maximization, Cluster Evaluation Techniques – Internal and External, clustering ensemble.
Linear Discriminant Analysis, Support Vector Machine, Naïve Bayes, Gradient Descent, Class Evaluation Measures, Overfitting, Bias Variance Trade-off Precision, Recall, F1 Score, ROC, AUC, Validation Strategies.
Neural Network, LTU, Perceptron, MLP, Activation Function, Loss Functions, Optimizers, Momentum Adadelta, RMSProp, Adam, Early Stopping, drop-out, Batch Normalization; Word Embedding, CboW, Skip-gram, Glove, ElMo, CNN, RNN, LSTM, GRU, Encoder-Decoder Network, Transfer, Auto Encoder, Generative Adversarial Network; Ensemble Methods - Bagging, Committee Machines and Stacking, Ensemble Methods – Boosting, Gradient Boosting; Undirected Graphical Models, Markov Chains, Random walk Monte Carlo, HMM, Variable elimination, belief propagation,

Introduction to Reinforcement Learning, Sampling-based techniques, Q Learning, Introduction to XAI, Lime, SHAP etc. Learning from online streaming data, Machine Learning Applications.

Books:

1. Machine Learning, Tom Mitchell, McGraw Hill, 1997.
2. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. Feb 2009. Trevor Hastie, Robert Tibshirani, Jerome Friedman.
3. Introduction to Machine Learning, third edition. Ethem Alpaydin. The MIT Press. September 2014: ISBN: 978-0-262-028189

MDS-SEC–302. Big Data Algorithms & Analytics
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction, Dawn of the Big Data Era, Definition and Features of Big Data, Big Data Value, The Development of Big Data, Challenges of Big Data, Big Data and its importance, Drivers, Big data analytics, Storage System for Massive Data, Distributed Storage System, Storage Mechanism for Big Data - Database Technology, Design Factors, Database Programming Model
Sketching, Streaming, dimensionality reduction, randomized algorithms, universal hashing, Markov chains and random walks, sampling methods.
Traditional Data Analysis, Big Data Analytic Methods, Architecture for Big Data Analysis - Real-Time vs. Offline Analysis, Analysis at Different Levels, Analysis with Different Complexity, Tools for Big Data Mining and Analysis.
Data Storage and Analysis, Comparison with Other Systems, A Brief History of Hadoop, Apache Hadoop and the Hadoop Ecosystem, Analyzing the Data with Unix Tools, Analyzing the Data with Hadoop(MapReduce, Java MapReduce), Scaling Out, Hadoop Streaming, Hadoop Pipes, Task trackers, Hadoop Configuration, NoSQL Data Management, Hadoop with R and Python, Introduction to Apache Spark, Hadoop vs Spark, Map-reduce using Spark, Big Data tools and techniques, Pig and Hive.
Applications of big data analytics in healthcare, transportation, finance and banking, IoT and sensor networks, social networks, NLP, smart cities etc.

Text Books:

1. Boris lublinsky, Kevin T. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”, Wiley, ISBN: 9788126551071, 2015.
2. Chris Eaton, Dirk Deroos et al. “Understanding Big data”, McGraw Hill,2012.
3. Tom White, “HADOOP: The definitive Guide”, O Reilly2012.
4. Karau,Holden,etal.*Learningspark:lightning-fastbigdataanalysis*."O'ReillyMedia,Inc.", 2015.
5. Prajapati, Vignesh. *Big data analytics with R and Hadoop*. Packt Publishing Ltd,2013.
6. Radtka, Zachary, and Donald Miner. *Hadoop with Python*. O'Reilly Media,2015.
7. Cloud Computing Principles and Paradigms – RajkumarBuyya.
8. Distributed Systems - GeorgeCoulouris.
9. Cloud Application Architectures - GeorgeReese

MDS-DSE–303. Elective - II
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

MDS-DSE–304. Elective - III
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

MDS-COR–311. Machine Learning & Deep Learning Lab
Full Marks: 100, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P
Lab pertaining to MDS-COR-301

MDS-SEC–312. Big Data Algorithms & Analytics Lab
Full Marks: 100, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P
Lab pertaining to MDS-SEC-302

MDS–DSE-312. Advanced Programming Laboratory
Full Marks: 100, Weekly Hours: 0 + 0 + 3
Allotted Hrs: 40P

MDS-DISS–321. Dissertation-I
Full Marks: 100, Weekly Hours: 0 + 0 + 3

Semester-IV

MDS-DISS-421. Dissertation-II
Full Marks: 300

MDS-GV-431. Grand viva: Each student will face a grand viva.
Full Marks: 100

Department Specific Elective (DSE) Papers

Elective: Time Series Analysis
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Overview of forecasting. Models for time series: Time-dependent seasonal components. Autoregressive (AR), moving average (MA) and mixed ARMA-modeller. The Random Walk Model. Box-Jenkins methodology. Forecasts with ARIMA and VAR models.

Dynamic models with time-shifted explanatory variables. The Koyck transformation . Partial

adjustment and adaptive expectation models. Granger's causality tests. Stationarity, unit roots and cointegration.Modelling of volatility: ARCH - and the GARCH-models.

Elective: Multicriteria Decision Making
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Different Multicriteria Decision Making (MCDM) algorithms for Big Data Analysis, Importance, Features, Case-based Studies, TOPSIS, Cognitive Map, VIKOR, SEM, Entropy, Fuzzy cognitive map etc. Applications in Retail, Supply Chain, Logistic Sector using MCDM approaches. Optimization in Routing, Behavioral, Model formulation in Decision making problems based on Data Sciences.
Real-life Applications – Multi-Echelon Supply Chain, railway management, Environment Concerning consumer behavior.

Elective: Speech &Natural Language Processing
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction Regular Languages and Finite State Automata, Morphology with Finite State Transducers, Text Categorization using Naive Bayes, Sentiment Mining, Log Linear Models, Language Models, POS Tagging with Hidden Markov Models, Models for Sequential tagging – MaxEnt, CRF, Syntax– Constituency Parsing, Dependency Parsing, Distributional Semantics, Lexical Semantics, Topic Models, Entity Linking, Named Entity Recognition with MEMMs, Brown Clustering, Conditional Random Fields for NER and POS Tagging, Information Extraction, Information Extraction, Document Similarity in Information Retrieval, Statistical Natural Language Parsing, Other NLPTasks

Elective: Computation Biology
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction to molecular biology, cell, chromosome, DNA, RNA, proteins, Central Dogma, protein structures, computational biology and bioinformatics tasks;
Sequence databases, sequence comparison, sequence alignment, local and global sequence alignment, multiple sequence alignment, web tools for sequence comparisons;
Sequencing, genome sequencing, fragment assembly, next-generation sequencing, handling errors in sequencing, gene finding, promoter identification, sequence-based protein classification;
Protein structures, structure prediction from sequence, motif finding, structure alignment, structure-based protein classification, molecular design and docking;
Phylogeny analysis, phylogenetic tree construction algorithms, parsimony and distance-based techniques;
Gene expression analysis, microarray, microarray analysis, differential expression, microarray clustering,biclustering, classification, gene marker prediction, gene selection, gene ordering, gene prioritization, gene significance analysis, gene co-expression, differential co-expression;
Biological networks, protein-protein interactions, gene regulatory networks, metabolic networks, network analysis and prediction, systems biology;
Biological databases, sequence databases, gene/protein databases, protein structure/domain databases, microarray gene expression databases, protein-protein interaction databases, gene regulatory network

databases, metabolic network databases.

Text Books:

1. Carlos Setubal and Joao Meidanis, "Introduction to Computational Molecular Biology", Brooks/Cole.

Reference Books:

1. Molecular Cell Biology by DaidBaltimar
2. Aurther M. Lesk, Introduction to Bioinformatics, Oxford University Press, 4th edition (2014)
3. Dan E. Krane and Michael L. Raymer, Fundamental Concepts of Bioinformatics Krane and Raymer, DORLING KINDERSLEY (RS); First edition (2003)
4. David Mount : Bioinformatics: Sequence and Genome Analysis, CBS; 2 edition (2005)

Elective: Information Security
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

The Security Problem in Computing: The meaning of computer Security, Computer Criminals, Methods of Defense, Elementary Cryptography: Substitution Ciphers, Transpositions, Making "Good" Encryption algorithms, The Data Encryption Standard, The AES Encryption Algorithms, Public Key Encryptions, Uses of Encryption.

Program Security: Secure Programs, Nonmalicious Program Errors, viruses and other malicious code, Targeted Malicious code, controls Against Program Threats, Protection in General-Purpose operating system protected objects and methods of protection memory and addmens protection, File protection Mechanisms, User Authentication Designing Trusted O.S: Security polices, models of security, trusted

O.S design, Assurance in trusted O.S. Implementationexamples.

Database Security: Security requirements, Reliability and integrity, Sensitive data, Inference, multilevel database, proposals for multilevel security. Security in Network: Threats in Network, Network Security Controls, Firewalls, Intrusion Detection Systems,SecureE-Mail.

Administering Security: Security Planning, Risk Analysis, Organizational Security policies, Physical Security. Legal Privacy and Ethical Issues in Computer Security: Protecting Programs and data, Information and the law, Rights of Employees and Employers, Software failures, Computer Crime, Praia, Ethical issues in Computer Security, case studies ofEthics

Elective: Mobile Applications & Programming
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Elective: Cloud Computing
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction to cloud computing – Overview of Computing, Cloud Computing NIST Model, Properties, characteristics and disadvantages, role of open standards.

Cloud computing architecture – cloud computing stack, service Models (XaaS), IaaS, Paas, SaaS, Daas,

Deployment Models, private, public, hybrid, commercial cloud models.
Service management in Cloud computing – service level agreement (SLA), SLA violation, cloud economics.
Resource management in cloud computing – resource sharing, scalability, elasticity, transparency.
Data management in cloud computing – looking at data scalability and cloud services, database and data stores in cloud, large scale data processing
Cloud security – infrastructure security, data security and storage, identity and access management, access control, trust, reputation risk
Cloud simulators – CloudSim, CloudAnalyst, MultiRecCloudSim, CloudSimPlus, GreenCloudSimulator
Research trend in Cloud compting, green cloud computing, fog computing

Text Books:

1. Cloud Computing Bible by Barrie Sosinsky, Wiley India Pvt. Ltd, 2013
2. Mastering Cloud Computing by RajkumarBuyya, Christian Vecchiola, S. ThamaraiSelvi, McGraw Hill
3. Education (India) Private Limited, 2013

Reference Books:

1. Cloud computing: A practical approach, Anthony T. Velte, Tata Mcgraw-Hill
2. Cloud Computing, Miller, Pearson
3. Building applications in cloud:Concept, Patterns and Projects, Moyer, Pearson

Elective: Digital Image Processing and Computer Vision
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Image Formation and Coordinate Transformations, Camera Matrix, Motion/Stereo Pin-hole model, Human eye / cognitive aspects of colour / 3D space; illumination; Sampling and Quantization Coordinate transformations and cameraparameters
Image Processing - Noise Removal, Blurring, Edge Detection: Canny / Gaussian/ Gabor/ Texture Edges/Curvature / Corner Detection. Motion Estimation: Horn-Schunk Optical Flow Formulation Euler-Lagrange formulation: Calculus of variations theory. Structure Recovery from Motion Segmentation - Concept of Figure vs. Ground, Watershed, Change Detection, Background Subtraction, Texture Segmentation, Gaussian Mixture Models - Applications in Color/Motion based ImageSegmentation, Background Modeling and Shape Clustering
Machine Learning techniques in Vision, Bayesian Classification, Maximum Likelihood Methods, Neural Networks; Non-parametric models; Manifold estimation, Support Vector Machines; Temporal sequence learning
Introduction to Object Tracking - Exhaustive vs. Stochastic Search, Shapes, Contours, and Appearance Models, Mean-shift tracking; Contour-based models
Object Modeling and Recognition, Fundamental matrix / Epipolar geometry, Adaboost approaches: Face Detection / Recognition, Large Datasets; Attention models.
Applications: Surveillance, Object detection, etc.

Text Book:

1. Digital Image Processing, Gonzalves,Pearson
2. Digital Image Processing, Jahne, Springer India
3.Digital Image Processing &Analysis,Chanda&Majumder,PHI
4.Fundamentals of Digital Image Processing, Jain, PHI

References Books:
1.Image Processing, Analysis & Machine Vision, Sonka, VIKAS
2. Getting Started with GIS- Clarke Keith. C; PE.
3. Concepts & Techniques of GIS - Lo C.P, Albert, Yeung K.W- PHI.

Elective: Web-based Programming
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Elective: Soft Computing
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L
Introduction: Introduction to soft computing; introduction to biological and artificial neural network; introduction to fuzzy sets and fuzzy logic systems.
Introduction to Genetic Algorithm, Genetic Operators and Parameters, Genetic Algorithms in Problem Solving, Theoretical Foundations of Genetic Algorithms, Implementation Issues.
Artificial neural networks and applications: Different artificial neural network models; learning in artificial neural networks; neural network applications in control systems. Neural Nets and applications of Neural Network.
Fuzzy systems and applications: fuzzy sets; fuzzy reasoning; fuzzy inference systems; fuzzy control; fuzzy clustering; applications of fuzzy systems.
Neuro-fuzzy systems: neuro-fuzzy modeling; neuro-fuzzy control.
Applications: Pattern Recognitions, Image Processing, Biological Sequence Alignment and Drug Design, Robotics and Sensors, Information Retrieval Systems, Share Market Analysis, Natural Language Processing.

Text Books:
1. M. Mitchell: An Introduction to Genetic Algorithms, Prentice-Hall.
2. J.S.R.Jang, C.T.Sun and E.Mizutani: Neuro-Fuzzy and Soft Computing, PHI, Pearson Education.
3. Timothy J.Ross: Fuzzy Logic with Engineering Applications, McGraw-Hill.
4. Davis E.Goldberg: Genetic Algorithms: Search, Optimization and Machine Learning, Addison Wesley.

Reference Books:
1. S. Rajasekaran and G.A.V.Pai: Neural Networks, Fuzzy Logic and Genetic Algorithms, PHI.
2. D. E. Goldberg: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley.

Elective: Complex Network Analysis
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L
Introduction, Overview of Network science, Motivation, Large scale dynamic networks, Challenges of graph theory Basic Concepts related to Networks Small world effect, transitivity and clustering,

degree distribution, scale free networks, maximum degree; network resilience; mixing patterns; degree correlations; community structures; network navigation Community Structure Analysis Basic concepts of network communities, Modularity, various community finding approaches like Girvan-Newman Algorithm, Spectral Bisection Algorithm, Radicchi Edge Clustering Algorithm (for binary as well as weighted graphs), Wu-Hubermann Algorithm, and Random Walk based Algorithm, Louvain, InfoMap Random Graphs Poisson random graphs, generalized random graphs, the configuration model,generating functions, power-law degree distribution, directed graph, bipartite graph, degree correlations Models of Network Growth Price model, Barabasi& Albert model, other growth models, vertex copying models, Bipartite Network Processes taking place on Networks Percolation theory and network resilience, Epidemiological processes, Cascades and information spread Social Network Homophily, Cohesiveness, Cliques, Clans, Clubs, Plex, Equivalence of ties, Ego-centric networks, Cascade formation and information diffusion in Social media (say Twitter). Applications, Search on networks, exhaustive network search, guided network search, network navigation; network visualization and semantic zooming. Advanced topics, Temporal network, Multilayer networks, Interdependent networks, Controllability of complex networks, Economic and financial network analytics

Elective: Operations Research
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction, Mathematical Modeling, Linear programming Formulation, solution procedures, Duality, Sensitivity, Applications, Network methods Max Flow, Min cost, Shortest path, Dynamic programming Sequential decisions, Principle of optimality, Applications Integer Programming Formulation, Nonlinear Programming. Applications and solution methods.

Elective: High Performance Computing
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Parallel Processing Concepts (Quick Overview): Levels of parallelism (instruction, transaction, task, thread, memory, function). Models (SIMD, MIMD, SIMT, SPMD, Dataflow Models, Demand-driven Computation etc). Architectures: N-wide superscalar architectures, multi-core, multi-threaded

Parallel Programming with CUDA: Processor Architecture, Interconnect, Communication, Memory Organization, and Programming Models in high performance computing architectures: (Examples: IBM CELL BE, Nvidia Tesla GPU, Intel Larrabee Microarchitecture and Intel Nehalem microarchitecture). Memory hierarchy and transaction specific memory design. Thread Organization

Fundamental Design Issues in Parallel Computing: Synchronization. Scheduling. Job Allocation. Job Partitioning. Dependency Analysis. Mapping Parallel Algorithms onto Parallel Architectures. Performance Analysis of Parallel Algorithms.

Fundamental Limitations Facing Parallel Computing: Bandwidth Limitations. Latency Limitations. Latency Hiding/Tolerating Techniques and their limitations

Power-Aware Computing and Communication: Power-aware Processing Techniques. Power-aware

Memory Design. Power-aware Interconnect Design. Software Power Management

Advanced Topics: Petascale Computing. Optics in Parallel Computing. Quantum Computers. Recent developments in Nanotechnology and its impact on HPC

Reference Books:

1. "Highly Parallel Computing", by George S. Almasi and Alan Gottlieb
2. "Advanced Computer Architecture: Parallelism, Scalability, Programmability", by Kai Hwang, McGraw Hill 1993.
3. "Parallel Computer Architecture: A hardware/Software Approach", by David Culler Jaswinder Pal Singh, Morgan Kaufmann, 1999.
4. "Scalable Parallel Computing", by Kai Hwang, McGraw Hill 1998.
5. "Principles and Practices on Interconnection Networks", by William James Dally and Brian Towles, Morgan Kauffman 2004.
6. GPU Gems 3 --- by Hubert Nguyen (Chapter 29 to Chapter 41).
7. Introduction to Parallel Computing, Ananth Grama, Anshul Gupta, George Karypis, and Vipin Kumar, 2nd edition, Addison-Welsey, © 2003.
8. Petascale Computing: Algorithms and Applications, David A. Bader (Ed.), Chapman & Hall/CRC Computational Science Series, © 2007.

Elective: Blockchain Technology
Full Marks: 100, Weekly Hours: 3 + 1 + 0
Allotted Hrs: 40L

Introduction of Cryptography and Blockchain: What is Blockchain, Blockchain Technology Mechanisms & Networks, Blockchain Origins, Objective of Blockchain, Blockchain Challenges, Transactions And Blocks, P2P Systems, Keys As Identity, Digital Signatures, Hashing, and public key cryptosystems, private vs. public Blockchain.
BitCoin and Cryptocurrency: What is Bitcoin, The Bitcoin Network, The Bitcoin Mining Process, Mining Developments, Bitcoin Wallets, Decentralization and Hard Forks, Ethereum Virtual Machine (EVM), Merkle Tree, Double-Spend Problem, Blockchain And Digital Currency, Transactional Blocks, Impact Of Blockchain Technology On Cryptocurrency.
Introduction to Ethereum: What is Ethereum, Introduction to Ethereum, Consensus Mechanisms, How Smart Contracts Work, Metamask Setup, Ethereum Accounts, Receiving Ether's What's a Transaction?, Smart Contracts.
Introduction to Hyperledger: What is Hyperledger? Distributed Ledger Technology & its Challenges, Hyperledger & Distributed Ledger Technology, Hyperledger Fabric, Hyperledger Composer.
Solidity Programming:
Solidity - Language of Smart Contracts, Installing Solidity & Ethereum Wallet, Basics of Solidity, Layout of a Solidity Source File & Structure of Smart Contracts, General Value Types (Int, Real, String, Bytes, Arrays, Mapping, Enum, address)
Blockchain Applications: Internet of Things, Medical Record Management System, Domain Name Service and Future of Blockchain, Alt Coins.

Reference Books:

1. Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller and Steven Goldfeder, Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction, Princeton University Press (July 19, 2016).
2. Antonopoulos, Mastering Bitcoin.
3. Antonopoulos and G. Wood, Mastering Ethereum.
4. D. Drescher, Blockchain Basics. Apress, 2017.