DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

PG : SCHEME AND SYLLABUS (2023-25)

MTech in DATA SCIENCE

PG Curriculum Course Credits

2023-25

Scheme Structure of M.Tech Program from the Academic year 2023-25

Semester	Professional Core Courses (PCC)	Professional Core Courses Lab (PCL)	Professional Courses - Electives (PEC)	Project Work / Seminar (PW)	Internship (INT)	Total semester load
First	6	2	11	-	-	22
Second	8	2	12	-	-	22
Third	4	-	4	4	4	16
Fourth	-	-	-	20	-	20
Total	**18**	**4**	**27**	**24**	**4**	**80**

M.Tech. in Data Science
Scheme of Teaching and Examination 2023-25

I SEMESTER

Sl. No.	Subject Code	Subject	Teaching Department	Category	Credits				Total contact hours / week
					L	T	P	Total	
1	MDS11	Data Management	ISE	PCC	3	0	0	3	3
2	MDS12	Statistics for Data Science	ISE	PCC	2	1	0	3	4
3	MDSE13x	Professional Core Elective	ISE	PCE	3	0	0	3	3
4	MDSE14x	Professional Core Elective	ISE	PCE	3	1	0	4	5
5	MDSE15x	Professional Core Elective	ISE	PCE	4	0	0	4	4
6	RM116	Research Methodology and IPR	ISE	MCC	3	0	0	3	3
7	MDSL16	Data Visualization Lab	ISE	PCL	0	0	1	1	2
8	MDSL17	Big data lab	ISE	PCL	0	0	1	1	2
Total					**19**	**1**	**2**	**22**	**26**

MDSE13x	MDSE14x	MDSE15x
Artificial intelligence	Linear algebra and optimization	Cloud Computing
Advanced Software Engineering	Time series analysis and Forecasting	Scalable Data Systems
Advanced algorithms	Optimization for Data Analysis	Distributed Storage Technologies

M.Tech. in Data Science
Scheme of Teaching and Examination 2023-25

II SEMESTER

Sl. No.	Subject Code	Subject	Teaching Department	Category	Credits				Total contact hours /week
					L	T	P	Total	
1	MDS21	Machine Learning	ISE	PCC	3	1	0	4	5
2	MDS22	Deep Learning and Reinforcement Learning	ISE	PCC	4	0	0	4	4
3	MDSE23x	Professional Core Elective	ISE	PCE	4	0	0	4	4
4	MDSE24x	Professional Core Elective	ISE	PCE	4	0	0	4	4
5	MDSE25x	Professional Core Elective	ISE	PCE	4	0	0	4	4
6	MDSL26	Machine Learning lab	ISE	PCL	0	0	1	1	2
7	MDSL27	Deep Learning Lab	ISE	PCL	0	0	1	1	2
8	MDSMC	Current Trends in Industry	ISE	MC	0	0	0	0	1
Total					**19**	**1**	**2**	**22**	**26**

MDSE23x	MDSE24x	MDSE25x
Full Stack Development	Exploratory Data Analysis	Natural Language Processing
Blockchain Technology	Social Network Analysis	Bioinformatics
Data Engineering and MLOps	Computer Vision	Business Analytics

M.Tech. in Data Science
Scheme of Teaching and Examination 2023-25

III SEMESTER

Sl. No.	Subject Code	Subject	Teaching Department	Category	Credits				Total contact hours /week
					L	T	P	Total	
1	MDS31	Data Security and Privacy	ISE	PCC	3	1	0	4	5
2	MDSE32x	Professional Core Elective	ISE	PCE	4	0	0	4	4
3	MDSI33	Internship/Industrial Training	ISE	INT	0	0	4	4	8
4	MDSP34	Project Work - I	ISE	PW	0	0	4	4	8
Total					**7**	**1**	**8**	**16**	**25**

Where x=1,2,3…

L –Lecture, T – Tutorial, P- Practical

MDSE32x
Cognitive Computing
Explainable AI
IoT Analytics

M.Tech. in Data Science
Scheme of Teaching and Examination 2023-25

IV SEMESTER

Sl. No.	Subject Code	Subject	Teaching Department	Category	Credits				Total contact hours /week
					L	T	P	Total	
1	MDSP41	Project Work - II	ISE	PW	0	0	20	20	40
Total					**0**	**0**	**20**	**20**	**40**

L –Lecture, T – Tutorial, P- Practical

DATA MANAGEMENT	
Course Code: MDS11	**Credits: 3:0:0**
Pre – requisites: NIL	**Contact Hours: 42**
Course Coordinator: Dr Sumana M	

Course Contents

Unit 1

Principles of Data Management, SQL for Data Science: SQL Basics, SQL Joins and aggregates, Grouping and query evaluation, SQL Sub-queries, Key Principles of RDBMS, Entity Relationship models, design principles and constraints. The Enhanced Entity–Relationship (EER) Model.

Unit 2

High level Data Models- From E/R diagram to Relational Design, Object Definition Language Characteristics of Data Warehouses, Data Modeling for Data Warehouses , Building a Data Warehouse , Typical Functionality of a Data Warehouse , Data Warehouse versus Views .

Unit 3

Introduction to Transaction Processing, Transaction and System Concepts, Desirable Properties of Transactions, Characterizing Schedules Based on Recoverability, Characterizing Schedules Based on Serializability, Transaction Support in SQL. Two-Phase Locking Techniques for Concurrency, Concurrency Control Based on Timestamp Ordering, Multiversion Concurrency Control Techniques

Unit 4

Introduction to NOSQL Systems, Characteristics of NOSQL Systems, Categories of NOSQL Systems, The CAP theorem, MongoDB Data Model, MongoDB CRUD Operations, MongoDB Distributed Systems Characteristics, NOSQL Key-Value Stores, Replication, Sharding, Cassandra: Architecture, Data Model.

Unit 5

The Architecture of a Search Engine: Components of a Search Engine, Web Crawlers, Query Processing in Search Engines, Ranking Pages, PageRank for Identifying Important Pages: Recursive Formulation of PageRank, Spider Traps and Dead Ends, Topic-Specific PageRank: Calculating A Topic-Specific PageRank

Text Books:

1. Fundamentals of database systems by Elsmasri and Navathe, 7th Edition, 2015
2. Database Systems: The Complete Handbook, by Hector Garcia-Molina, Jennifer Widom, and Jeffrey Ullman. Second edition, 2008.

3. Seven NoSQL Databases in a Week: Get up and running with the fundamentals, By Xun (Brian) Wu, Sudarshan Kadambi, Devram Kandhare, Aaron Ploetz, Packt Publishers, 2018

STATISTICS FOR DATA SCIENCE	
Course Code: MDS12	**Credits: 2:1:0**
Pre – requisites: NIL	**Contact Hours: 28L+28T**
Course Coordinator: Dr. Krishna Raj	
Course Contents	
Unit 1	
Exploratory Data Analysis: estimates of locations and variability, exploring data distributions, exploring binary and categorical data, exploring two or more variables.	
Unit 2	
Random sampling and bias, selection bias, sampling distribution of statistic, bootstrap, confidence intervals, data distributions: normal, long tailed, student's-t, binomial, Chi-square, F, Poisson and related distributions.	
Unit 3	
A/B testing, hypothesis testing, resampling, statistical significance & p-values, t-tests, multiple testing, degrees of freedom.	
Unit 4	
ANOVA, chi-square test, multi-arm bandit algorithm, power and sample size, simple linear regression, multiple linear regression.	
Unit 5	
Prediction using regression, factor variables in regression, interpreting the regression equation, regression diagnostics, polynomial and spline regression **TUTORIALS** Numerical problems on the various statistical concepts to be covered in the tutorial sessions.	
Reference Books: 1. Peter Bruce, Andrew Bruce and Peter Gadeck, "Practical Statistics for Data Scientists", 2nd edition, O'Reilly Publications, 2020. 2. Allan B Downey, "Think Stats", Version 1.6.0, Green Tea Press, 2011.	

ARTIFICIAL INTELLIGENCE	
Course Code: MDSE131	**Credits: 3:0:0**
Pre – requisites: NIL	**Contact Hours: 42**
Course Coordinator: Dr. Rajeshwari S B	

Course Contents

Unit 1

Introduction: Definition of AI. Foundation of Artificial Intelligence. **Intelligent Agents:** Agents and Environments, Rationality, The Nature of Environments, The Structure of Agents.
Problem-solving by search: Problem-Solving Agents, Uninformed Search Strategies: Bidirectional Search. Informed Search Strategies: A* Search, Heuristic Functions.

Unit 2

Adversarial Search: Games, Optimal Decisions in Games, Alpha Beta Pruning, Imperfect Real-Time Decision. **Logical Agents:** Knowledge-Based Agents, The Wumpus World, Logic,
Propositional Logic: A very simple logic, Effective Propositional Model Checking, Agents Based on Propositional Logic. **First Order Logic:** Wumpus World representation, Knowledge Engineering in First-Order Logic.

Unit 3

Interference in First-order Logic: Propositional vs. First-Order Inference, Unification and Lifting, Forward chaining, Backward Chaining. Resolution. **Classical Planning:** Definition, Algorithms for Planning as State-Space Search, Planning Graphs, Other Planning Approaches.

Unit 4

Knowledge Representation: Ontological Engineering, Categories and Objects, Events, Mental Objects and Modal Logic, Reasoning Systems for Categories
Uncertainty: Acting under Uncertainty, Inference using Full Joint Distributions, Independence

Unit 5

Uncertainty: The Wumpus World Revisited, Learning from Examples: Forms of Learning.
Robotics: Introduction, Hardware, Perception, Planning to Move, Planning Uncertain Movement, Moving, Robotic Software Architecture, Application Domains.

Text Books:

1. Stuart Russel, Peter Norvig: Artificial Intelligence - A Modern Approach, 3rd Edition, Pearson Education, 2012.
2. Stephen Marsland, "Machine Learning - An Algorithmic Perspective", Second Edition, CRC Press - Taylor and Francis Group, 2015

Reference Books:

1 Peter Jackson, "Introduction to Expert Systems", 3rd Edition, Pearson Education, 2007.
2 Deepak Khemani "Artificial Intelligence", Tata Mc Graw Hill Education 2013.
3 http://nptel.ac.in

ADVANCED SOFTWARE ENGINEERING	
Course Code: MDSE132	**Credits: 3:0:0**
Pre – requisites: NIL	**Contact Hours: 42**
Course Coordinator: Dr. Krishnaraj P M	
Course Contents	
Unit 1	
Evolution of Software Engineering, Process Models– Waterfall Model, incremental model, Spiral Model, prototyping model, Agile Process and Principles, Scrum, Test-driven development, Rational Unified Process, maturity models, Critical Analysis of Process Models.	
Unit 2	
DevOps – Introduction, DevOps life cycle, Principles, benefits, Roles, Responsibilities and skills for a DevOps engineer, DevOps versus Agile, Continuous Integration and Deployment, Tools at various stages of DevOps. MLOps - Processes and Tools. Requirements Engineering - Product versus service requirements engineering. Functional and Non-functional requirements, Requirements engineering process.	
Unit 3	
Software Architecture Models, Software Design – Concepts, Design principles, Object Oriented Design with UML, Software Design Patterns, Universal Design in Software Engineering. User Interface Design – Introduction, Golden rules, interface analysis and design process, Interface design steps, WebApp interface design	
Unit 4	
Programming Paradigms – Imperative programming, Functional programming, Logical programming, Object oriented programming, Global software development – tools and practices, Coding standards. Software Quality Assurance models. Free and Open-Source Software Engineering.	
Unit 5	
Software Configuration Management - Processes and Tools, Software Project Management - estimation of time, personnel and budget, task allocation. IEEE standards in Software Engineering. Software Engineering for Data Science- Process, Standards and Tools.	
Text Books: 1. Roger S Pressman, Software Engineering, 7th edition, TMH publication. 2. Ian Sommerville, Software Engineering, 9th edition, Pearson Education. 3. Rumbaugh, Object –Oriented Modeling and Design, Pearson Education.	
Reference Books: 1. Course Pack prepared by faculty members.	

ADVANCED ALGORITHMS	
Course Code: MDSE133	**Credits: 3:0:0**
Pre – requisites: NIL	**Contact Hours: 42**
Course Coordinator: Karthik V	
Course Contents	
Unit 1	
Analysis techniques: Growth of functions: Asymptotic notation, Standard notations and common functions, Substitution method for solving recurrences, Recursion tree method for solving recurrences, Master theorem. **Amortized Analysis**: Aggregate analysis, the accounting method, the potential method.	
Unit 2	
Sorting in Linear Time: Lower bounds for sorting, Counting sort, Radix sort, Bucket sort. **Advanced Design and Analysis Technique**: Matrix-chain multiplication, longest common subsequence, Elements of the greedy strategy, an activity-selection problem	
Unit 3	
Graph Algorithms: Bellman-Ford Algorithm, Shortest paths in a DAG, Johnson's Algorithm for sparse graphs. **Maximum Flow**: Flow networks, Ford Fulkerson method and Maximum Bipartite Matching	
Unit 4	
Number Theoretic Algorithms: Elementary notions, GCD, Modular arithmetic, solving modular linear equations, The Chinese remainder theorem, Powers of an element, RSA cryptosystem	
Unit 5	
Advanced Data structures: Structure of Fibonacci heaps, Mergeable-heap operations, decreasing a key and deleting a node, Binomial Queues, Splay Trees. **String Matching Algorithms**: Naïve algorithm, Rabin-Karp algorithm, String matching with finite automata, Knuth-Morris-Pratt algorithm.	
Text Books: 1. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein; Introduction to Algorithms; Columbia University. 2. Mark Allen Weiss; Data Structures and Algorithm Analysis in C++, Addison-Wesley.	
Reference Books: 1. Kozen DC, The design and analysis of algorithms, Springer Science & Business Media. 2. Kenneth A. Berman, Jerome L. Paul,Algorithms, Cengage Learning.	

LINEAR ALGEBRA AND OPTMIZATION	
Course Code: MDSE141	**Credits: 3:1:0**
Pre – requisites: NIL	**Contact Hours: 42L+14T**
Course Coordinator: Dr Monica Anand	
Course Contents	
Unit 1	
Matrices and Linear Equations: Matrices, Vectors and Linear Equations, Geometric interpretation of solution of linear equations, Inverse matrices, Gauss Jordan method, Factorization A = LU **Vector Spaces:** Interpreting vectors as data, Linear combinations of vectors, Linear independence of vectors, Vector spaces, Basis and dimension, Null space and column space of A.	
Unit 2	
Linear Transformations: Matrices of linear transformation, Rotation, Reflection, Dilation and contraction, Co-ordinate vectors and change of basis **Projections and Least Squares Approximation:** Orthogonality, Projections, Orthonormal bases and Gram-Schmidt method, Factorization A = QR, Least squares approximations	
Unit 3	
Singular Value Decomposition: Eigenvalues and eigenvectors, Symmetric matrices, Diagonalization of symmetric matrices and interpretation as basis expansion, Quadratic forms, Positive definite and positive semidefinite matrices, Singular Value Decomposition (SVD) and Interpretation as basis expansion, Principal Component Analysis (PCA)	
Unit 4	
Vector Calculus: A Machine Learning View, Introduction, The Basics of Optimization, Univariate Optimization, Differentiation of univariate functions, Partial differentiation and gradients, Gradients of vector valued functions, gradient of matrices, Backpropagation and automatic differentiation, Higher order derivatives, Linearization and multivariate Taylor's series, Constrained optimization and Lagrangian multipliers, Hessian matrix	
Unit 5	
Optimization: Loss functions, Objective function, Unconstrained optimization: Gradient descent, Checking Gradient Correctness with Finite Differences, Optimization Models for Binary Targets - Least-Squares Classification: Regression on Binary Targets, Why Least-Squares Classification Loss Needs Repair, Weighted least squares and regularized least squares, Support Vector Machine	
Text Books: 1. "Linear Algebra and Optimization for Machine Learning A Textbook" Aggarwal, Charu, Springer International Publishing Print ISBN: 978-3-030-40343-0 2. Mathematics for Machine Learning: Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Cambridge University Press, 2020	
Reference Books: 1. Basics of Linear Algebra for Machine Learning, Jason Brownlee, Online free download Book.	

2. Introductory Linear Algebra, an Applied First Course, Bernard Kolman and David R.Hill, Pearson, Eighth Edition, 2009.
3. Convex Optimization, Stephen Boyd and Lieven Vandenberghe, Cambridge University Press, First Edition, 2018

TIME SERIES ANALYSIS AND FORECASTING	
Course Code: MDSE142	**Credits: 3:1:0**
Pre – requisites: NIL	**Contact Hours: 42L + 14T**
Course Coordinator: Savita K Shetty	
Course Contents	
ptimization: Loss functions, Objective function, Unconstrained optimization: Gradient descen hecking Gradient Correctness with Finite Differences, Optimization Models for Binary Target Least-Squares Classification: Regression on Binary Targets, Why Least-Squares Classificatio oss Needs Repair, Weighted least squares and regularized least squares, Support Vecto lachine	
Unit 1	
Time Series Decomposition and Smoothing : Trend and Seasonal Components of Time Series Exponential Smoothing of Time Series Forecasting with Exponential Smoothing	
Unit 2	
Least Squares Prediction Forecasting with Classical Regression Models Forecast Error Statistics and Evaluation	
Unit 3	
Time Series Analysis I: Introduction Covariance Stationarity Trend in Time Series Unit Root Problem: Estimation and Testing	
Unit 4	
Time Series Analysis II: ARIMA Models ,Identification , Estimation and Diagnostic Checking Forecasting, ARMA Analysis of Regression Residuals	
Unit 5	
Time Series Analysis III: Advanced Topics :ARCH and GARCH Model Estimation, Multi-Equation Time Series Models, Dynamic Linear Models	
Text Books: 1. Paul S.P. Cowpertwait and Andrew V. Metcalfe, Introductory Time Series with R, Springer-Verlag, New York, 2009. 2. Rob J. Hyndman and George Athanasopoulos, Forecasting: Principles and Practice, One line, Open Access Textbooks.	
Reference Books: 1. Walter Zucchini, Oleg Nenadic, Time Series Analysis with R.	

2. Avril Coghlan, A Little Book of R for Time Series.

OPTIMIZATION FOR DATA ANALYSIS	
Course Code: MDSE143	**Credits: 3:1:0**
Pre – requisites: NIL	**Contact Hours: 42L + 14T**
Course Coordinator: Dr Manisekhar R S	
Course Contents	
Unit 1	
Data Analysis and Optimization, Least Squares, Matrix Factorization Problems, Solutions to Optimization Problems: Taylor's Theorem , Characterizing Minima of Smooth Functions, Convex Sets and Functions, Strongly Convex Functions, Descent Directions	
Unit 2	
Steepest-Descent Method: General Case, Convex Case, Strongly Convex Case. Descent Methods: Convergence, Line-Search Methods(Direction and Steplength), Convergence to Approximate Second-Order Necessary Points, Mirror Descent, The KL and PL Properties	
Unit 3	
Gradient Methods Using Momentum: Motivation from Differential Equations, Nesterov's Method: Convex Quadratics, Convergence for Strongly Convex Functions, Convergence for Weakly Convex Functions, Conjugate Gradient Methods, Lower Bounds on Convergence Rates. Stochastic Gradient: Noisy Gradients	
Unit 4	
Stochastic Gradient: Incremental Gradient Method, Classification and the Perceptron, Empirical Risk Minimization, Randomness and Step length: Insights, Example: Computing a Mean, The Randomized Kaczmarz Method, Key Assumptions for Convergence Analysis, Case : Bounded Gradients: Lg = 0, Convergence Analysis : Case : Lg = 0, Implementation Aspects : Epochs ,Mini batching	
Unit 5	
First-Order Methods for Constrained Optimization: Optimality Conditions Euclidean Projection, The Projected Gradient Algorithm, General Case: A Short-Step Approach, General Case: Backtracking Nonsmooth Functions and Subgradients: Subgradients and Subdifferentials, The Subdifferential and Directional Derivatives, Calculus of Subdifferentials, Convex Sets and Convex Constrained Optimization	
Text Books: Stephen J. Wright, Benjamin Recht. Optimization for Data Analysis,Cambridge University Press, 2022	

Reference Books:

Panos M. Pardalos, Athanasios Migdalas, Open Problems in Optimization and Data Analysis, Springer Nature 2018

CLOUD COMPUTING	
Course Code: MDSE151	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr Manisekhar R S	

Course Contents

Unit 1

Introduction: Network centric computing and network centric content, Peer-to-peer systems, Cloud Computing, Cloud Computing delivery models & Services, Ethical issues, Cloud vulnerabilities, Challenges. **Cloud Infrastructure:** Amazon, Google, Azure & online services, open source private clouds. Storage diversity and vendor lock-in, intercloud, Energy use & ecological impact of data centers, service level and compliance level agreement, Responsibility sharing, user experience, Software licensing.

Unit 2

Cloud Computing: Applications & Paradigms, Challenges, existing and new application opportunities, Architectural styles of cloud applications, Workflows: Coordination of multiple activities, Coordination based on a state machine model – the ZooKeeper, The MapReduce programming model, A case study: the Grep TheWeb application, Clouds for science and engineering, High performance computing on a cloud, cloud computing for biological research, Social computing, digital content, and cloud computing.

Unit 3

Cloud Resource Virtualization: Virtualization, Layering and virtualization, Virtual machine monitors, Virtual machines, Performance and security isolation, Full virtualization and paravirtualization, Hardware support for virtualization, Case study: *Xen* -a VMM based on paravirtualization, Optimization of network virtualization in *Xen* 2.0, *vBlades* -paravirtualization targeting a *x86-64* Itanium processor, A performance comparison of virtual machines, The darker side of virtualization, Software fault isolation.

Unit 4

Cloud Resource Management and Scheduling: Policies and mechanisms for resource management, Applications of control theory to task scheduling on a cloud, Stability of a two-level resource allocation architecture, Feedback control based on dynamic thresholds, Resource bundling, combinatorial auctions for cloud resources, Scheduling algorithms for computing clouds, fair queuing, Cloud scheduling subject to deadlines.

Unit 5

Storage systems: Storage models, file systems, databases, DFS, General parallel File system, GFS, Apache Hadoop, Locks & Chubby, TPS & NOSQL databases, Bigdata, Mega store. **Cloud security:** Risks, Security, privacy and privacy impacts assessments, Trust, VM Security, Security of virtualization, Security risks in shared images.

Text Books:

1. Dan Marinescu, Cloud Computing: Theory and Practice, 1st edition, MK Publishers, 2013.

Reference Books:

1. Kai Hwang, Jack Dongarra, Geoffrey Fox, Distributed and Cloud Computing, From Parallel Processing to the Internet of Things, 1st edition, MK Publishers, 2012.

2 Anthony T. Velte, Toby J. Velete, Robert Elsenpeter, Cloud Computing: A Practical Approach, Tata McGraw Hill, 2010.

SCALABLE DATA SYSTEMS	
Course Code: MDSE152	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr Sanjay H A	

Course Contents

Unit 1

Distributed System Models and Enabling Technologies: Scalable Computing Over the Internet, Technologies for Network-Based Systems, System Models for Distributed and Cloud Computing, Software Environments for Distributed Systems and Clouds, Performance, Security, and Energy Efficiency

Unit 2

MapReduce Basics : Functional Programming , Mappers and Reducers , The Execution Framework , Partitioners and Combiners , The Distributed File System , Hadoop Cluster Architecture .
Basic MapReduce Algorithm Design: Local Aggregation , Pairs and Stripes

Unit 3

Big Data Processing with MapReduce and Spark, Spark Basics, RDD, transformations, action, Shuffle, Spark internals & Spark tuning

Unit 4

Introduction to Kafka: Messages and Batches , Schemas, Topics and Partitions , Producers and Consumers , Brokers and Clusters , Multiple Clusters, Benefits of Kafka, Use cases,
Configuring Kafka, Programming with Kafka: Producers and Consumers

Unit 5

Introduction : Need for Kubernetes, Container Technology, Introduction to Kubernetes
Creating, running, and sharing a container image :Building the container image, Running the container image, Exploring the inside of a running container , Stopping and removing a container , Pushing the image to an image registry,
Setting up an alias and command-line completion for kubectl ,
Pods: running containers in Kubernetes : Introducing pods , Creating a simple YAML descriptor for a pod, Organizing pods with labels

Text Books:

1. Kai Hwang, G. C. Fox, J.J. Dongarra "Distributed & Cloud Computing", Morgan Kauffman Publishers
2. Learning Spark, Holden Karau, et al., 1st Edition
3. Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale
4. Data-Intensive Text Processing, with MapReduce , Jimmy Lin and Chris Dyer
5. Kubernetes in Action, MARKO LUKŠA

DISTRIBUTED STORAGE TECHNOLOGIES	
Course Code: MDSE153	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Shashidhara H S	
Course Contents	
Unit 1	
Introduction: Information Storage, Evolution of Storage Architecture, Data Centre Infrastructure, Virtualization and Cloud Computing. **Data Centre Environment**: Application, DBMS, Host, Connectivity, Storage, Disk Drive Components, Disk Drive Performance, Host Access to Data, Direct-Attached Storage, Storage Design Based on Application, Disk Native Command Queuing, Introduction to Flash Drives.	
Unit 2	
Data Protection: RAID Implementation Methods, Array Components, Techniques, Levels, Impact on Disk Performance, Comparison, Hot Spares. **Intelligent Storage System:** Components, Storage Provisioning, Types.	
Unit 3	
Fibre Channel Storage Area Networks: FC Overview, Evolution, Components, FC Connectivity, Ports, FC Architecture, Fabric Services, Login Types, Zoning, FC Topologies, Virtualization in SAN. **IP SAN and FCoE:** iSCSI, FCIP, FCoE.	
Unit 4	
Network-Attached Storage: Benefits, Components, NAS I/O Operation, Implementations, File Sharing Protocols, Factors Affecting NAS Performance, File-Level Virtualization. **Object Based and Unified Storage:** Object Based Storage Devices, Content Addressed Storage, CAS Use Cases, Unified Storage.	
Unit 5	
Business Continuity: Information Availability, Terminology, Planning Lifecycle, Failure Analysis, Impact Analysis, Solutions. **Cloud Computing:** Cloud Enabling Technologies, Characteristics, Benefits, Service Models, Deployment Models, Infrastructure, Challenges, Adoption Considerations. **Securing the Storage Infrastructure**: Framework, Risk Triad, Domains. **Managing the Storage Infrastructure:** Monitoring, Management Activities, Management Challenges, Information Lifecycle Management, Storage Tiering.	
Text Books:	

Somasundaram G, Alok Shrivastava, (EMC Education Services), Information Storage and Management, 2e, Wiley India, 2012, ISBN 9788126537501

Reference Books:

1. Robert Spalding; Storage Networks: The Complete Reference, Tata McGraw Hill, 2003.
https://education.emc.com/ISMbookv2/resources_content.aspx

DATA VISUALIZATION LAB	
Course Code: MDSL16	**Credits: 0:0:1**
Pre – requisites: NIL	**Contact Hours: 28**
Course Coordinator: Savita Shetty	

Course Contents

1. Understanding the association between two continuous or quantitative factors
   - Scatterplots
   - Correlation

2. Understanding the association between two continuous or quantitative factors
   - Simple Linear Regression
   - F-test for Simple Linear Regression
   - t-test for Simple Linear Regression

3. Regression diagnostics
   - Residual Plots
   - Outliers and Influence Points Assumptions of least-square regression

4. Multiple linear regression
   - Equation of multiple linear regression
   - Interpretation of multiple linear regression
   - F-test for Multiple Linear Regression
   - t-tests in Multiple Linear Regression
   - Cautions about Regression

5. Analysis of Variance (ANOVA)
   - One-Way Analysis of Variance
   - F-test for ANOVA
   - Evaluating Group Differences
   - Type I and Type II Errors

6. Analysis of Variance (ANOVA)
   - Issues with Multiple Comparisons
   - Assumptions of Analysis of Variance
   - Relationship between One-Way Analysis of Variance and Regression
   - One-Way Analysis of Covariance
   - Two-Way Analysis of Variance
   - Two-Way Analysis of Covariance

7. Analysis for proportions
   - One-Sample Tests for Proportions
   - Significance Tests for a Proportion
   - Confidence Intervals for a Proportion

8. Analysis for proportions
   - Two-Sample Tests for Proportions
   - Confidence Intervals for Differences in Proportions
   - Significance Tests for Differences in Proportions
   - Effect Measures
   - Logistic Regression
   - Multiple Logistic Regression
   - Area under the ROC Curve

Part A: Experiments performed using tools such as Tableau and Power BI
Part B: Experiments performed in Python.

Text Books:

1. Andy Field, Jeremy Miles and Zoe Field. (2012) Discovering Statistics Using R. Publisher: SAGE Publications Ltd. ISBN-13: 978-1446200469
2. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. (2013) An Introduction to Statistical Learning with Applications in R. Springer
3. https://www.openintro.org/stat/ Free PDF for download & R tutorials and codes.

Reference Books:

1. Teetor, P. (2011). R cookbook. Sebastopol, CA: O'Reilly. ISBN 9780596809157.
2. Chang, W. (2013). R graphics cookbook. Sebastopol, CA: O'Reilly. ISBN 9781449316952

BIG DATA LAB	
Course Code: MDSL17	**Credits: 0:0:1**
Pre – requisites: NIL	**Contact Hours: 28**
Course Coordinator: Kusuma S	

Course Contents

1. Exploring few Analytical tools: MS Excel, SAS(http://www.sas.com/en_us/hOome.html), Statistica(http://www.statsoft.com/) and open source Analytical tool: R Analytics(http://www.revolutionalanalytics.com)
2. Perform setting up and Installing Hadoop in its three operating modes: Standalone, Pseudo distributed, fully distributed.
3. Implement the following file management tasks in Hadoop:
    1. Adding files and directories
    2. Retrieving files
    3. Deleting files Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities.
4. Implement the following Queries using MongoDB Querying language.
    1. Create the collection by name: Student" and insert documents.
    2. Demonstrate save method
    3. Add new field to the existing document
    4. Find the document based on Search Criteria
    5. Demonstrate Count, Limit and Skip methods
5. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.
6. Write a Map Reduce program that mines weather data. (Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with Map Reduce, since it is semi structured and record-oriented).
7. Implement Matrix Multiplication with Hadoop Map Reduce paradigm.
8. Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data.
9. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.
10. Solve some real life big data problems.

Text Books:

Big Data Analytics, Seema Acharya and Subhashini Chellappan. Wiley India Pvt. Ltd. 2015.

Reference Books:

1. Arshdeep Bahga, Vijay Madisetti, "Big Data Analytics: A Hands-On Approach", 1st Edition, VPT Publications, 2018. ISBN-13: 978-0996025577.
2. Raj Kamal and Preeti Saxena, "Big Data Analytics Introduction to Hadoop, Spark, and Machine Learning", McGraw Hill Education, 2018 ISBN: 9789353164966, 9353164966

3. S C Albright and W L Winston," Business analytics: data analysis and decision making", 5/e Cengage Learning.

II SEMESTER

MACHINE LEARNING	
Course Code:MDS21	**Credits: 3:1:0**
Pre – requisites: NIL	**Contact Hours: 42L+28T**
Course Coordinator: Dr Sumana M	
Course Contents	
Unit 1	
Text Data: Flattening, Filtering and chunking: Bag-of-X: Turning Natural Text into Flat Vectors, Filtering for cleaner features, Atoms of Meaning: From words to n-Grams to Phrases. **Categorical variables:** Encoding categorical variables, dealing with large categorical variables: feature hashing, Bin counting **Dimensionality reduction:** Intuition, Derivation, PCA in Action, Whitening and ZCA, Considerations and limitations of PCA, Singular value decomposition, PCA as special case for SVD	
Unit 2	
Regression Analysis: Linear Regression, Multiple Linear Regression, Logistic Regression, Hypothesis space and logistic regression, Bias-Variance trade-offs in Regression, Case study **Decision Tree Learning:** Decision Trees- Basic algorithm (ID3), Hypothesis search and Inductive bias, Entropy and Gain calculations, Issues in Decision Tree Learning – Overfitting, Solutions to overfitting, Dealing with continuous values	
Unit 3	
Decision Tree Improving performance: Bagging and Boosting, Adaboost - combining weak learners, Adaboost - simple problems. Random Forests: Methods for Growing the Trees, choose m attributes randomly, compute their information gains, and choose the attribute with the largest gain to split, Generalization Error of Random Forests, Random Forest Regression **Supervised Learning** - K-Nearest Neighbour Technique, Examples as Numerical	
Unit 4	
Support Vector Machines: Linear and Nonlinear - Technique, Examples as Numerical; Kernel Functions, Dimensionality Reduction - Supervised: **Bayesian Learning:** Bayes theorem – An Example; Bayes theorem and concept learning: Brute-Force Bayes Concept Learning, MAP Learning and Consistent Learners, Maximum Likelihood Estimation, Bayes optimal classifier, Naive Bayes classifier, Bayesian Belief Network- Conditional Independence, Representation, Inference, Learning Bayesian Belief Networks	
Unit 5	
Un-supervised Learning: Hierarchical vs non-hierarchical clustering, Agglomerative and divisive clustering, Expectation Maximization, Gaussian Mixtures, EM Clustering, K-means clustering, Simple problems, Bisecting k- means, issues with k-means. K Means as special case of Expectation Maximization	
Reference Books:	

1. Stephen Marsland, "Machine Learning - An Algorithmic Perspective", Second Edition, CRC Press - Taylor and Francis Group, 2015.
2. Ethem Alpaydin, "Introduction to Machine Learning", Second Edition, MITPress, Prentice Hall of India (PHI) Learning Pvt. Ltd. 2010.
3. Tom M. Mitchell, "Machine Learning", McGraw-Hill Education, ISBN: 978-1-25-909695-2, 2013.
4. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining" Pearson, 4th edition, ISBN: 978-81-717-1472-0, 2009.
5. Corrina cortes, Valdimir Vapnik, "Support Vector Networks" Kluwer Academic Publishers, 1995.

DEEP LEARNING & REINFORCEMENT LEARNING	
Course Code: MDS22	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr Mani Sekhar	
Course Contents	
Unit 1	
Foundations of Neural Networks and Deep Learning: Neural Networks, The Biological Neuron, The Perceptron, Multilayer Feed-Forward Networks, Training Neural Networks, Backpropagation Learning. **Fundamentals of Deep Networks (DN):** Architectural Principles of Deep Networks, Building Blocks of Deep Networks. **Major Architectures of Deep Networks:** Unsupervised Pretrained Networks, Deep Belief Networks, Generative Adversarial Networks,	
Unit 2	
Recurrent Neural Networks: Modeling the Time Dimension, 3D Volumetric Input, Markov Models, General Recurrent Neural Network Architecture, Long short-term memory (LSTM) Networks, **Building Deep Networks:** Matching Deep Networks to the Right Problem, Columnar Data and Multilayer Perceptions	
Unit 3	
Building Deep Networks: Images and Convolutional Neural Networks, Time-series Sequences and Recurrent Neural Networks, Hybrid Networks. **Tuning Deep Networks**: Basic Concepts in Tuning Deep Networks, Building Deep Networks, and Step-by-Step Process in building the Intuition, Matching Input Data and Network Architectures, Feed-Forward Multilayer Neural Networks	
Unit 4	

Reinforcement Learning: Introduction RL, Elements of RL, limitations & scope, Multi armed bandits: k-armed bandit, action value method, 10-armed testbed, incremental implementation, tracking nonstationary problem, gradient bandit algorithm **Finite Markov Decision process:** agent environment interface, goals & rewards, returns & episodes, notation for episodic & continuing task, policies and value function
Unit 5
Dynamic programming: Policy- evaluation, improvement, iteration, value iteration, Asynchronous dynamic programming, generalized policy iteration **Monte Carlo method**: prediction, estimation of action values, control, off policy prediction via important sampling. **Temporal difference learning**: Temporal difference prediction, advantages of TD, optimization of TD, SARSA, Q-learning
Text Books: 1. Josh Patterson & Adam Gibson, Deep Learning – A Practitioners Approach, O'Reilly, 1st Edition 2017. 2. Richard S and Andrew G. Reinforcement learning – An introduction, The MIT Press Cambridge, Massachusetts, 2020. http://incompleteideas.net/book/RLbook2020.pdf
Reference Books: 1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, An MIT Press book, http://www.deeplearningbook.org 2. http://cse.iitkgp.ac.in/~aritrah/course/theory/RL/Autumn2022/

FULL STACK DEVELOPMENT	
Course Code: MDSE231	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr. Sumana M	
Course Contents	
Unit 1	
Introduction to Bootstrap : Bootstrap Basics , Bootstrap Grids , Bootstrap Themes , Bootstrap CSS , Bootstrap JS, Introduction to JavaScript , JavaScript Language Basics , JavaScript Objects , JavaScript Scope , JavaScript Events , JavaScript Strings, JavaScript Numbers , JavaScript Math , JavaScript Arrays , JavaScript Boolean , JavaScript Comparisons , JavaScript Conditions, JavaScript Switch , JavaScript Loops , JavaScript Type Conversion , JavaScript RegExp, JavaScript Errors , JavaScript Debugging , JavaScript Hoisting , JavaScript Strict Mode, JavaScript Functions , JavaScript Objects , JavaScript Forms , JavaScript HTML DOM	
Unit 2	

Angular (Type Script) , Introduction to Angular , Angular Application Architecture , What is NgModule , Angular Components , Angular Templates , Data Binding , Types of Data Binding , Modules Component Working , Directives , Structure Directives , Template Routing , Theme Implementation in Angular Framework , Angular Forms , Services , Inject Services , Angular Server Communication With Backend Server , Working of Api's(GET,POST,PUT,DELETE) , Complete Web application In Angular Framework
Unit 3
Nodejs , Introduction to Nodejs , Architecture of Nodejs Application , Synchronous and Asynchronous Programming , Call back Function in nodejs , Promises in Nodejs , Mongodb with Nodejs , Design the Schema in Nodejs , Design the Rest API's , GET,POST,PUT,DELETE , JSON web Token Authentication in nodejs , Create the Auth APP in nodejs , Create the E-commerce Backend , Integrated Payment Gateway
Unit 4
ExpressJS: Restful services, Introducing Express Building your First Web Server, Nodemon , Environment Variables , Route Parameters , Handling HTTP GET Request , Handling HTTP POST Request , Calling Endpoints Using Postman , Input Validations , Handling HTTP PUT Request , Handling HTTP DELETE Request , Project- Build the Genres API. Express- Advanced Topics, Middleware, Creating a Custom Middleware, Built-in Middleware , Environments , Configuration , Debugging , Templating Engine , Database Engines , Database Integration , Authentication , Structuring Express Applications
Unit 5
MongoDB, Introduction to MongoDB (No-sql), Collections in MongoDb , Documents In mongoDb , Difference between Mysql and NoSql , Inserting data into database , Filter queries in Mongodb Database , Schema Validation in MongoDb database , Indexing In collections , Aggregation in MongoDb , Embedded Document in MongoDb
Text Books: 1. Write Modern Web Apps with the MEAN Stack Mongo, Express, AngularJS, and Node.js DEVELOP AND DESIGN Jeff Dickey, 2019
Reference Books:

DATA ENGINEERING AND MLOps	
Course Code: MDSE232	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr Lincy Meera Mathews	
Course Contents	
Unit 1	
Data Engineering: Definition, The Data Engineering Lifecycle, Evolution of the Data Engineer, Data Engineering and Data Science, Data Engineering Skills and Activities, Data Maturity and the Data	

Engineer, The Background and Skills of a Data Engineer, Business Responsibilities, Technical Responsibilities, The Continuum of Data Engineering Roles, Data Engineers Inside an Organization , Internal-Facing Versus External-Facing Data Engineers, Data Engineers and Other Technical Roles, Data Engineers and Business Leadership. **Data Engineering Lifecycle:** The Data Lifecycle Versus the Data Engineering Lifecycle, Generation: Source Systems, Major Undercurrents Across the Data Engineering Lifecycle
Unit 2
Data Architecture: Enterprise Architecture Defined, Data Architecture Defined, "Good" Data Architecture, Principles of Good Data Architecture, Major Architecture Concepts, Domains and Services , Distributed Systems, Scalability, and Designing for Failure ,Tight Versus Loose Coupling: Tiers, Monoliths, and Microservices , User Access: Single Versus Multitenant , Event-Driven Architecture , Examples and Types of Data Architecture **Choosing Technologies Across the Data Engineering Lifecycle:** Team Size and Capabilities, Speed to Market, Interoperability, Cost Optimization and Business Value, Total Cost of Ownership Total Opportunity Cost of Ownership, FinOps, Today Versus the Future: Immutable Versus Transitory Technologies: Hybrid Cloud, Multicloud , Decentralized: Blockchain and the Edge ,Monolith Versus Modular , Serverless Versus Servers, Server Versus Serverless evaluation
Unit 3
MLOps Challenges, MLOps to Mitigate Risk, Risk Assessment,Risk Mitigation, MLOps for Responsible AI,MLOps for Scale. **Key MLOps Features:** Model Development, Establishing Business Objectives, Data Sources and Exploratory Data Analysis, Feature Engineering and Selection, Training and Evaluation, Reproducibility, Responsible AI, Productionalization and Deployment, Model Deployment Types and Contents, Model Deployment Requirements, Monitoring **Developing Models**: Machine Learning Model, Required Components, Different ML Algorithms, Different MLOps Challenges, Data Exploration, Feature Engineering and Selection, Feature Engineering Techniques, How Feature Selection Impacts MLOps Strategy, Experimentation, Evaluating and Comparing Models, Choosing Evaluation Metrics, Cross-Checking Model Behavior, Impact of Responsible AI on Modeling, Version Management and Reproducibility
Unit 4
Preparing for Production: Runtime Environments, Adaptation from Development to Production Environments, Data Access Before Validation and Launch to Production, Final Thoughts on Runtime Environments, Model Risk Evaluation, The Purpose of Model Validation, The Origins of ML Model Risk, Quality Assurance for Machine Learning **Deploying to Production:** CI/CD Pipelines, Building ML Artifacts, The Testing Pipeline, Deployment Strategies, Categories of Model Deployment, Considerations When Sending Models to Production, Maintenance in Production, Containerization, Scaling Deployments, Requirements and Challenges
Unit 5

Monitoring and Feedback Loop: Models Be Retrained, Understanding Model Degradation, Ground Truth Evaluation, Input Drift Detection, Drift Detection in Practice, Example Causes of Data Drift, Input Drift Detection Techniques, The Feedback Loop, Logging, Model Evaluation, Online Evaluation

Model Governance: Governance the Organization Needs, Matching Governance with Risk Level, Current Regulations Driving MLOps Governance, Pharmaceutical Regulation in the US: GxP Financial Model Risk Management Regulation, GDPR and CCPA Data Privacy Regulations, The New Wave of AI-Specific Regulation, The Emergence of Responsible AI, Key Elements of Responsible AI (Element 1 to element 5), A Template for MLOps Governance (Step 1 to 8)

Text Books:

1. Joe Reis, Matt Housley Fundamentals of Data Engineering: Plan and Build Robust Data Systems, O'Reilly, 2022
2. Mark Treveil and the Dataiku Team. Introducing MLOps How to Scale Machine Learning in the Enterprise, O'Reilly, 2020

BLOCKCHAIN TECHNOLOGY	
Course Code: MDSE233	**Credits: 4:0:0**
Pre – requisites: Fundamentals of Distributed System	**Contact Hours: 56**
Course Coordinator: Dr Sanjay H A	

Course Contents

Unit 1

Distributed systems, CAP theorem, Byzantine Generals problem, Consensus. The history of blockchain, Introduction to blockchain, Various technical definitions of blockchains, Generic elements of a blockchain, Features of a blockchain, Applications of blockchain technology, Tiers of blockchain technology, Consensus in blockchain, CAP theorem and blockchain, Benefits and limitations of blockchain

Unit 2

Decentralization using blockchain, Methods of decentralization, Blockchain and full ecosystem decentralization, Smart contract, Decentralized organizations, Decentralized autonomous organizations, Decentralized autonomous corporations, Decentralized autonomous societies Decentralized applications, Platforms for decentralization,

Cryptographic primitives: Symmetric cryptography, Asymmetric cryptography, Public and private keys Hash functions: Compression of arbitrary messages into fixed length digest, Easy to compute, Pre-image resistance, Second pre-image resistance, Collision resistance, Message Digest (MD), Secure Hash Algorithms (SHAs), Merkle trees, Patricia trees, Distributed hash tables (DHTs), Digital signatures, Elliptic Curve Digital signature algorithm (ECDSA)

Unit 3

Bitcoin, Bitcoin definition, Transactions, The transaction life cycle, The transaction structure, Types of transaction, The structure of a block, The structure of a block header, The genesis block, The bitcoin network, Wallets, Smart Contracts-History, Definition, Ricardian contracts, Smart contract templates, Oracles, Smart Oracles, Deploying smart contracts on a blockchain, The DAO
Unit 4
Ethereum 101, Introduction, Ethereum clients and releases, The Ethereum stack, Ethereum blockchain, Currency (ETH and ETC), Forks, Gas, The consensus mechanism, The world state, Transactions, Contract creation transaction, Message call transaction, Elements of the Ethereum blockchain , Ethereum virtual machine (EVM), Accounts, Block, Ether, Messages, Mining, The Ethereum network Hands-on: Clients and wallets -Get
Unit 5
Hyperledger, Hyperledger as a protocol, Fabric, Hyperledger Fabric, Sawtooth lake, Corda
Text Books: Imran Bashir. “Mastering BlockChain”, Packt
Reference Books: Mastering Bitcoin: Programming the Open Blockchain Paperback – 2017 by Andreas M.O’rielly

EXPLORATORY DATA ANALYSIS	
Course Code: MDSE241	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr. Pushpalatha M N	
Course Contents	
Unit 1	
Introduction to Exploratory data analysis and Ipython: IPython: Beyond Normal Python, IPython Magic Commands, Input and Output History, IPython and Shell Commands, Errors and Debugging, Understanding Data Types in Python, Aggregations: Min, Max, and Everything in Between, Computation on Arrays: Broadcasting, Comparisons, Masks, and Boolean Logic, Fancy Indexing, Sorting Arrays, Structured Data: NumPy's Structured Arrays	
Unit 2	
Introduction to Pandas and Data Manipulations: Introducing Pandas Objects, Data Indexing and Selection, operating on Data in Pandas, Combining Datasets: Merge and Join, Aggregation and Grouping, Pivot Tables, Vectorized String Operations, Working with Time Series, Dates and Times in Python, Pandas Time Series: Indexing by Time, Pandas Time Series Data Structures	
Unit 3	
Visualization with Matplotlib: Basics of simple plotting, Line Chart vs Line Graph, Bar Graph, Pie Chart, Histogram, Frequency Polygons, Box Plot, Scatter Plot, Saving Plots or Graph or chart to a file. General Matplotlib Tips, Two Interfaces for the Price of One, Visualizing Errors, Density and Contour Plots, Binnings, and Density, Customizing Plot Legends, Customizing Color bars, Multiple Subplots, Three-Dimensional Plotting in Matplotlib, Visualization with Seaborn.	
Unit 4	
Visualizations using Scikit-Learn for Machine Learning: Introducing Scikit-Learn, Hyper parameters and Model Validation, Feature Engineering, In Depth: Naive Bayes Classification, Bayesian Classification, Gaussian Naive Bayes, when to Use Naive Bayes, In Depth: Linear Regression, Simple Linear Regression, Basis Function Regression, Regularization, In Depth: Principal Component Analysis, PCA as Noise Filtering, Example: Eigenfaces	
Unit 5	
Visualizations using Scikit-Learn for Machine Learning II: In Depth: k-Means Clustering, k-Means Algorithm: Expectation–Maximization, In Depth: Gaussian Mixture Models, Motivating GMM: Weaknesses of k-Means, Generalizing E–M: Gaussian Mixture Models, GMM as Density Estimation, In-Depth: Decision Trees and Random Forests, Motivating Random Forests: Decision Trees, Ensembles of Estimators: Random Forests	
Text Books: 1. Python Data Science Handbook: Essential Tools for Working with Data, by Jake VanderPlas **2.** https://python4csip.com/files/download/Data%20Visualization.pdf	
Reference Books:	

1. Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt.
2. Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications, Glenn J. Myatt, and Wayne P. Johnson. Print ISBN:9780470222805 |Online ISBN:9780470417409 |DOI:10.1002/9780470417409

SOCIAL NETWORK ANALYSIS	
Course Code: MDSE242	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr. Anitha P	
Course Contents	
Unit 1	
Network Basics: Why Model Networks? A set of examples. Basic Network Concepts, Adjacency Matrices, Graphs, Notation, Nodes, and Links, Good Will Hunting Problems, Formal and Informal Networks. Centrality measures- What is "Centrality" and Why do we Study It? calculating Nodal Centrality Measures, Directed Networks and Centrality Measures, Location in the Network. Graph Level Measures- Density, Diameter, Centralization, Average Centralities, Network Topology.	
Unit 2	
Social Theory: Social Links- Individual Actors, Social Exchange Theory, Social Forces, Graph Structure, Agent Optimization Strategies in Networks. Subgroup Analysis: Subgroups, Organizational Theory, Random Groups, Heuristics for Subgroup Identification, Analysis Methods. Diffusion and Influence: Applications for Social Diffusion, Strain Theory.	
Unit 3	
Meta-Networks and Relational Algebra: Modes of Data, Source, Target, Direction, Multimode Networks, Bridging a Meta-Network, Strength of Ties. Sources of Data: Network Sampling, Measuring Links, Data Quality, Additional Ethnographic Data Collection Methods.	
Unit 4	
Information Networks and the World Wide Web: The Structure of the web, Link Analysis and Web Search, Sponsored Search Markets - Encouraging Truthful Bidding in Matching Markets: The VCG Principle.	
Unit 5	
Network Dynamics: Structural Models - Cascading Behavior in Networks, The Small-World Phenomenon, Epidemics - The SIR Epidemic model, The SIS epidemic model.	
Text Books:	
1. Armstrong, H., McCulloh, I., & Johnson, A. (2013). Social network analysis with applications. John Wiley & Sons. 2. Easley, D., & Kleinberg, J. (2010). Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge university press. 3. Jackson, M. O. (2008). Social and economic networks (Vol. 3). Princeton: Princeton university press.	

COMPUTER VISION	
Course Code: MDSE243	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Shruthi G	

Course Contents

Unit 1

Introduction to Computer Vision: Image Formation and Radiometry, Geometric Transformation, Geometric Camera Models, Image Reconstruction from a Series of Projections

Unit 2

Image Processing Concepts: Fundamentals of Image Processing, Image Transforms, Image Filtering, Color Image Processing, Mathematical Morphology, Image Segmentation

Unit 3

Image Descriptors and Features: Texture Descriptors, Colour Feature, Edge Detection, Object Boundary and Shape Representations, Interest or Corner Point Detectors, Histogram of Oriented Gradients (HOG), Scale Invariant Feature Transform (SIFT), Speeded up Robust Features (SURF), Saliency

Unit 4

Recognition: Fundamental Pattern Recognition Concepts: Introduction to Pattern Recognition, Linear Regression, Basic Concepts of Decision Functions, Elementary Statistical Decision Theory, Gaussian Classifier, Parameter Estimation, Clustering for Knowledge Representation, Dimension Reduction, Template Matching, Artificial Neural Network (ANN) for Pattern Classification, Convolutional Neural Networks (CNNs), Autoencoder

Unit 5

Applications of Computer Vision: Machine Learning Algorithms and their Applications in Medical Image Segmentation, Motion Estimation and Object Tracking, Face and Facial Expression Recognition, Gesture Recognition, Image Fusion, Programming Examples

Text Books:

1. Computer Vision and Image Processing, Fundamentals and Applications, By Manas Kamal Bhuyan
2. Computer Vision: Algorithms and Applications, Richard Szeliski
3. Handbook of Image and video processing by A L Bovik

Reference Books:

1. Image processing and computer vision by Milan Sonka
2. Digital image processing by Chanda and Dutta Majumdar

NATURAL LANGUAGE PROCESSING	
Course Code: MDSE251	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr. Rajeshwari S B	
Course Contents	
Unit 1	
NLP, A Primer: NLP in the Real World, NLP Tasks, What Is Language?, Building Blocks of Language, Why Is NLP Challenging?, Machine Learning, Deep, Learning, and NLP: An Overview, Approaches to NLP, Heuristics-Based NLP, Machine Learning for NLP, Deep Learning for NLP, Why Deep Learning Is Not Yet the Silver Bullet for NLP, NLP Pipeline.	
Unit 2	
Text Representation: Vector Space Models, Basic Vectorization Approaches, One-Hot Encoding, Bag of Words, Bag of N-Grams, TF-IDF, Distributed, Representations, Word Embeddings, Going Beyond Words, Distributed, Representations Beyond Words and Characters, Universal Text Representations, Visualizing Embeddings, Handcrafted Feature Representations	
Unit 3	
Text Classification: Applications, A Pipeline for Building Text Classification Systems, One Pipeline, Many Classifiers, Using Neural Embeddings in Text Classification, Deep Learning for Text Classification, Interpreting Text Classification Models, Learning with No or Less Data and Adapting to New Domains	
Unit 4	
Information Extraction: IE Applications, IE Tasks, The General Pipeline for IE, Keyphrase Extraction, Implementing KPE, Practical Advice, Named Entity Recognition, Building an NER System, NER Using an Existing Library, NER Using Active Learning, Practical Advice, Named Entity Disambiguation and Linking, NEL Using Azure API, Relationship Extraction, Approaches to RE, RE with the Watson API, Other Advanced IE Tasks, Temporal Information Extraction, Event Extraction, Template Filling, Chatbots	
Unit 5	
BERT: Starting Off with BERT, A Primer on Transformers, Understanding the BERT Model, Getting Hands-On with BERT	
Text Books: **1.** Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems – 2020. **2.** Getting Started with Google BERT: Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, and Harshit Surana	
Reference Books: 1. Tanveer Sidiqui, U.S Tiwary, "Natural Language Processing & Information Retrieval", Oxford University Press, 2008 2. Anne Kao & Stephen R Poteel, 'Natural Language & Text Mining", Springer- Verlag	

BIOINFORMATICS	
Course Code: MDSE252	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Shashidhara H S	
Course Contents	
Unit 1	
The genetic material, gene structure and information content, protein structure and function, chemical bonds, molecular biology tools	
Unit 2	
Dot plots, simple alignments, gaps, scoring matrices, the Needleman and Wunsch algorithm, semiglobal alignments, the Smith and Waterman algorithm, database searches – BLAST and FASTA	
Unit 3	
Patterns of substitutions within genes, estimating substitution numbers, molecular clocks Molecular phylogenetics, phylogenetic trees, distance matrix methods, maximum likelihood approaches	
Unit 4	
Parsimony, Inferred Ancestral Sequences, strategies for fast searches – branch and bound and heuristic searches, consensus trees, tree confidence, molecular phylogenies Genomics – 1: Prokaryotic genomes, prokaryotic gene structure, GC content and prokaryotic genomes, prokaryotic gene density, eukaryotic genomes	
Unit 5	
Genomics – 2: Eukaryotic gene structure Open reading frames, GC contents in eukaryotic genomes, gene expression, transposition, repetitive elements Amino acids, polypeptide composition, secondary structure, tertiary and quaternary structures, algorithms for modeling protein folding.	
Text Books: Dan E. Krane, Michael L. Raymer, Fundamental Concepts of Bioinformatics, Pearson Education, 2008	
Reference Books: 1. T K Attwood, D J Parry Smith, Introduction to Bioinformatics, Pearson Education,2004 2. Gary B. Fogel, David W. Corne, Evolutionary Computation in Bioinformatics, Morgan Kaufmann Publishers	

BUSINESS ANALYTICS	
Course Code: MDSE253	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Mushtaq Ahmed D M	

Course Contents

Unit 1

Introduction: Overview of the strategic impact of BAI across key industries-Analytics 3.0-the nature of analytical competition- what makes an analytical competitor- analytics and business performance- Competing on Analytics with Internal and external Processes- A Road Map to Enhanced Analytical Capabilities- Managing Analytical People- The Architecture of Business Intelligence –Essential Practice Skills for High-Impact Analytics Projects: Listening to client, Framing the central problem, Scoping a project, Defining metrics for success, Creating a work plan, Assembling data and expert sources.

Unit 2

Descriptive Analytics: Data Types and Scales, Types of Data Measurement Scales, Population and Sample, Measures of Central Tendency, Percentile, Decile, and Quartile, Measures of Variation, Measures of Shape −Skewness and Kurtosis, Data Visualization.

Unit 3

Data Analysis and Interpretation Process: Analysis and interpretation of data, Analysis of qualitative data, Process of data analysis, Types of analytical techniques in BI - Interview analysis, Utility of computer-aided research and procedures followed, Content analysis process, School Case Study.

Unit 4

Data Warehousing: Data Warehouse, Data Mart, Master Data Management, Dimension (Data Warehouse), Slowly Changing Dimension, Data Vault Modelling, Extract, Transform, Load, Star Schema, Mapping problems to machine learning tasks, Evaluating models, Validating models.

Unit 5

Essential Aspects of Business Intelligence: Context Analysis, Business Performance Management, Business Process Discovery, Information System, Organizational Intelligence, Data Visualization, Data Profiling, Data Cleansing, Process Mining, Competitive Intelligence, Operational Intelligence, Business Activity Monitoring, Complex Event Processing, Business Process Management, Metadata, Root Cause Analysis.

BUSINESS ANALYTICS

Process Mining, Competitive Intelligence, Operational Intelligence, Business Activity Monitoring, Complex Event Processing, Business Process Management, Metadata, Root Cause Analysis.

Text Books:

1. Sharda R, Delen D, Turban E, Aronson J, Liang T. P, (2014), Business Intelligence and Analytics: Systems for Decision Support, 10th edition, Pearson Education.
2. Drew Bentley, “Business Intelligence and Analytics” Library Press publication -2017
3. U. Dinesh kumar, “Business Analytical – The science of data driven decision making”, Wiley 2017

Reference Books:

1. Glenn J. Myatt, "Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining", John Wiley & Sons, Second Edition, 2014.
2. Carlo-Vercellis, "Business intelligence datamining and optimization for decision making", First Edition.
3. An Introduction to Business Analytics, Ger Koole, Lulu.com, 2019.

MACHINE LEARNING LAB	
Course Code: MDSL26	**Credits: 0:0:1**
Pre – requisites: NIL	**Contact Hours: 28**
Course Coordinator: Dr Lincy Meera Mathews	

Course Contents

1. Using Keras, develop a neural network model to classify the type of clothing using Fashion MNIST database.
   For any given binary class dataset, apply k nearest neighbor algorithm with at least 4 distance measure and analyze the performance metrics.
2. To train and test a binary decision using zoo data available at UCI Zoo Data Set Apply decision tree model, with at least three attributes selection measure and analyze the performance measure.
3. Implement linear regression with multiple variables to predict the prices of houses. Analyze the performance of XGBoost and multiple linear regression for the housing dataset.
4. Build a logistic regression model to predict whether a student gets admitted into a university. Apply Naïve Bayes Model for the above dataset and perform a comparative analysis.
5. To model a classifier for predicting whether a patient is suffering from any heart disease or not. Apply nonlinear SVM with different kernels on a given dataset and analyze the results.
6. For any given multiclass dataset and two class dataset, apply nearest neighbor algorithm with different values of k. Analyze the relationship between the number of neighbors and model performance.
7. Apply two clustering - AGNES and DIANA and analyze the model performance using appropriate performance measures.
8. Apply Random Forest Classifier on plant village dataset. Analyze the results on grey scale images and RGB images.
9. Apply any two synthetic data generation methods for any given imbalanced two class data and perform classification. Analyze with Adaboost model.
10. For a given binary class dataset, apply any traditional ML algorithm. Use two feature reduction techniques – Correlation coefficient and PCA, and LDA to reduce the dimensions of the dataset. Apply the same ML algorithm and analyze the results.
11. & 12 . Miniproject

Reference Books:

1. Stephen Marsland, "Machine Learning - An Algorithmic Perspective", Second Edition, CRC Press - Taylor and Francis Group, 2015.
2. Ethem Alpaydin, "Introduction to Machine Learning", Second Edition, MITPress, Prentice Hall of India (PHI) Learning Pvt. Ltd. 2010.
3. Tom M. Mitchell, "Machine Learning", McGraw-Hill Education, ISBN: 978-1-25-909695-2, 2013.
4. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining" Pearson, 4th edition, ISBN: 978-81-717-1472-0, 2009.
5. Corrina cortes, Valdimir Vapnik, "Support Vector Networks" Kluwer Academic Publishers, 1995

DEEP LEARNING LAB	
Course Code: MDSL27	**Credits: 0:0:1**
Pre – requisites: NIL	**Contact Hours: 28**
Course Coordinator: Dr Manisekhar R S	

Course Contents

Students must deploy the models listed below for a data analytics problem statement. Following that, conduct a performance analysis for the same:

i. Auto Encoder
ii. Long Short-Term Memory Networks
iii. Gated Recurrent Unit
iv. Deep Neural Network
v. Q-learning
vi. Extreme Gradient Boosting
vii. Adaptive Boosting

Text Books:

1. Josh Patterson & Adam Gibson, Deep Learning – A Practitioners Approach, O'Reilly, 1st Edition 2017.
2. Richard S and Andrew G. Reinforcement learning – An introduction, The MIT Press Cambridge, Massachusetts, 2020. http://incompleteideas.net/book/RLbook2020.pdf
3. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, An MIT Press book, http://www.deeplearningbook.org

CURRENT TRENDS IN INDUSTRY	
Course Code: MDSMC	**Credits: 1:0:0**
Pre – requisites: NIL	**Contact Hours: 14**
Course Coordinator: Prashanth Kambli	

Course Contents

Industry experts will deliver lectures on the following topics and expose students to the current trends in industry in the area of Data Science.

TOPICS:

1. Tableau for Data Visualization
2. AutoML

3. One Shot Learning / Few Shot Learning
4. Generative Models
5. Diffusion Models
6. Transfer Learning
7. Machine Learning on semi supervised or self-supervised data
8. Introduction to Large Language Models

III SEMESTER

DATA SECURITY AND PRIVACY	
Course Code: MDS31	**Credits: 3:1:0**
Pre – requisites: NIL	**Contact Hours: 42L+28T**
Course Coordinator: Dr Geetha V	
Course Contents	
Unit 1	
Computer Security Concepts, Intrusion Detection, Firewalls: Characteristics, Types. Classical Encryption Techniques Symmetric Cipher Model, Cryptography, Cryptanalysis and Brute-Force Attack, Substitution Techniques, Caesar Cipher, Monoalphabetic Cipher, Polyalphabetic Cipher, One Time Pad. Block Ciphers and the data encryption standard: Traditional block Cipher structure, stream Ciphers and block Ciphers, Motivation for the Feistel Cipher structure, the Feistel Cipher.	
Unit 2	
Principles of Public-key Cryptosystems, Public-Key Cryptosystems, Applications for Public-Key Cryptosystems, Requirements for Public-Key Cryptosystems. Public-Key Cryptanalysis. The RSA Algorithm, Description of the Algorithm, Computational Aspects, the Security of RSA. Other Public-Key Cryptosystems: Diffe-Hellman Key Exchange, The Algorithm, Key exchange protocols, Man-in-the-Middle Attack, Simple secret key distribution, Secret key distribution with confidentiality and authentication, A hybrid scheme. Public keys certificates, X.509 certificates. Public key infrastructure, PKIX Management Functions, PKIX Management Protocols.	
Unit 3	
Authentication Vs Authorization, Authentication Methods –Password authentication, Public Key Cryptography, Biometric authentication, Out of band, Authentication Protocols – SSL, Password Authentication Protocol (PAP), Kerberos, Email authentication,- PGP, Database authentication, Message authentication; secure hash functions and Authorization Approaches to hmac; public key cryptography principles; public-key cryptography algorithms, digital signatures, key management. Kerberos, x.509 directory authentication service. Authorization Definition, Multilayer authorization,	
Unit 4	
Understanding Privacy: Social Aspects of Privacy, Legal Aspects of Privacy and Privacy Regulations Effect of Database and Data Mining technologies on privacy challenges raised by new emerging technologies such RFID, biometrics, etc., Privacy Models Introduction to Anonymization, Anonymization models: K-anonymity, l-diversity, t-closeness, differential privacy, Database as a service	
Unit 5	
Using technology for preserving privacy. Statistical Database security, Inference Control, Inference Control with Semantic Web, Homomorphic Encryption, Secure Multi-party computation and Cryptography Privacy-preserving Data mining Hippocratic databases, Hippocratic databases privacy preservation. Emerging Applications: Social Network Privacy, Location Privacy, Query Log Privacy, Biomedical Privacy	
Text Books:	

1. Cryptography and Network Security Principles and Practice William Stallings, 6th edition, Pearson Education
2. The Algorithmic Foundations of Differential Privacy, Cynthia Dwork and Aaron Roth. DOI: 10.1561/0400000042.

Reference Books:

1. https://s3.amazonaws.com/assets.datacamp.com/production/course_6412/slides/chapter1. pdf
2. Privacy-Preserving Data Mining- Models and Algorithms, Charu C Aggarwal, Yu Philips, S., Springer
3. Principles of Information Security, Information Security Professional - Michael E. Whitman and Herbert J. Mattord,4th Edition, Thompson.

COGNITIVE COMPUTING	
Course Code: MDSE321	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr Vijayakumar B P	
Course Contents	
Unit 1	
Introduction: Foundation of cognitive computing, nn platforms- machine learning, reasoning, natural language processing, speech recognition and vision (object recognition), human–computer interaction, dialog and narrative generation among other technologies. Features: Adaptive, Interactive, Iterative, Stateful, context aware.	
Unit 2	
Design Principles for Cognitive Systems, Sentiment analysis, Natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications.	
Unit 3	
Cognitive analytics, Relationship between Big Data and cognitive computing, Principal benefit of utilizing cognitive analytics over traditional big data analytics. Representing knowledge in taxonomies and ontologies, Applying advanced analytics to cognitive computing, Using machine learning and deep learning neural networks to model cognition.	
Unit 4	
Role of cloud and distributed computing in cognitive computing, Business implications of cognitive computing, Cognitive analytics- Cognitive computing-technology platforms, Word processing documents, E-mails, videos, images, audio files, presentations, webpages, social media and many other data formats, tagging, analysis and insight generation.	
Unit 5	

Process of building a cognitive application, Building a cognitive healthcare application, Smarter cities: cognitive computing in government, Emerging cognitive computing areas, Future applications for cognitive computing

Reference Books:

1. Judith S. Hurwitz, Marcia Kaufman, Adrian Bowles, Cognitive Computing and Big Data Analytics, Wiley 2015.
2. Adnan Hashmi and Adnan Masood, Cognitive Computing Recipes: Artificial Intelligence Solutions Using Microsoft Cognitive Services and TensorFlow, APress, 2019.
3. Mark Watson, Introduction to Cognitive Computing, 2018, Copyright 2016-2018 Mark Watson.
4. Vijay V Raghavan, Venkat N. Gudivada, Venu Govindaraju, C.R. Rao, Cognitive Computing: Theory and Applications, Elsevier, 2016.
5. Peter Fingar, Vinton G. Cerf, Cognitive Computing: A Brief Guide for Game Changers, Meghan-Kiffer Press, 2015.

EXPLAINABLE ARTIFICIAL INTELLIGENCE	
Course Code: MDSE322	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Dr. Rajeshwari S B	

Course Contents

Unit 1

Introduction: Explainable AI- Definition and its Needs , Challenges ,Evaluation, Uses.
How LinkedIn Uses Explainable AI, PwC Uses Explainable AI for Auto Insurance Claims, Accenture Labs Explains Loan Decisions, DARPA Uses Explainable AI to Build "Third-Wave AI".
An Overview of Explainability: Explanations, Interpretability and Explainability. Explainability Consumers: Practitioners—Data Scientists and ML Engineers, Observers—Business Stakeholders and Regulators, End Users—Domain Experts and Affected Users.
Types of Explanations: Premodeling explainability, Intrinsic Versus Post Hoc Explainability, Local, Cohort, and Global Explanations, Attributions, Counterfactual, and Example-Based Explanations.

Unit 2

AI Model: The best way to optimize the interaction between human and AI, Forecasts are not necessarily useful, Criteria for evaluating explanations, Explainable to whom and why.
AI architecture: Fitness function for explainable AI, Deep neural network is great for explainable AI, The more multitasking the better, How to collect multitasking datasets, Neural network architecture.
Explainable AI methods, forms and frameworks: XAI methods and their classification: Based on the scope of expandability, Based on implementation, Based on applicability, Based on explanation level. Forms of explanation: Analytical explanation, Visual explanation, Rule-based explanation, Textual explanation. Framework for model interpretability and explanation: Explain like I'm 5, skater, Local-

interpretable model- agnostic explanations, Shapley additive explanations, Anchors, Deep learning important features.

Unit 3

Methods and metrics for XAI models: Bringing explainability to AI decision- need for explainable AI, Taxonomy of explaining AI decisions. Methods of XAI: Techniques of XAI, stages of AI explainability, Types of post-model explanation methods. Metrics for XAI: Evaluation metrics for XAI decisions. Use case: Explaining deep-learning models using Grad-Cam.

Evaluation Measures of XAI: Need for transparency and trust in AI, Black-box deep learning models. Classification of XAI methods: Global Vs local, Surrogate Vs visualization, Model specific Vs model agnostic, Pre-model Vs In-model Vs Post-model

Unit 4

XAI's evaluation methods: Mental model, Explanation usefulness and satisfaction, User trust and reliance, Human-AI task performance, Computational measures

XAI's explanation methods: Lime, sp-lime, DEEP-Lift, Layer-wise relevance propogation, Characteristics value evaluation, Reasoning from examples, Latent space traversal.

XAI stake holders: Developers, Theorists, Ethicists, Users.

XAI Applications: Training and tutoring, 6G, Network intrusion detection, Planning as a service, Prediction of non-communicable diseases, Scanning patients for COVID-19 signs

Unit 5

Hands-On Explainable AI (XAI) with Python: Open source XAI tools and techniques to process trustworthy AI results. Building models, Interpreting results with visualizations, and integrating XAI reporting tools and different applications.

Text Books:

1. Explainable AI: Foundations, Methodologies and Applications. Mayuri Mehta, Vasile Palade and Indranth Chatterjee, Springer, 2023.
2. Explainable AI for Practitioners, by Michael Munn, David Pitman, O'Reilly, 2022.
3. Hands-On Explainable AI (XAI) with Python, Denis Rothman, 2020

Reference Books:

1. Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning, John Liu, Uday Kamath, Springer, 2021.
2. Explainable AI Recipes, Pradeepta Mishra, Apress, 2023.

IOT ANALYTICS	
Course Code: MDSE323	**Credits: 4:0:0**
Pre – requisites: NIL	**Contact Hours: 56**
Course Coordinator: Jagadeesh Sai	

Course Contents
Unit 1
Defining IoT Analytics and Challenges, The situation, Defining IoT analytics, IoT analytics challenges, Business value concerns, IoT Devices and Networking Protocols, IoT devices, The wild world of IoT devices, Sensor types, Networking basics, IoT networking connectivity protocols, Connectivity protocols (when the available power is limited- Bluetooth Low Energy, 6LoWPAN, ZigBee, NFC, Sigfox), Connectivity protocols (when power is not a problem- Wi-Fi
Unit 2
IoT networking data messaging protocols, Message Queue Telemetry Transport (MQTT), Hyper-Text Transport Protocol (HTTP), Constrained Application Protocol (CoAP), Data Distribution Service (DDS), Hands-on on CoAP, MQTT, and HTTP, Websocket, Building elastic analytics, Elastic analytics concepts, Designing for scale, Cloud security and analytics, The AWS overview, AWS key concepts, AWS key core services, AWS key services for IoT analytics, The AWS CloudFormation overview, The AWS Virtual Private Cloud (VPC) setup walk-through, Creating a key pair for the NAT and bastion instances, Creating an S3 bucket to store data, Creating a VPC for IoT Analytics, How to terminate and clean up the environment, Collecting All That Data - Strategies and Techniques, Designing data processing for analytics
Unit 3
Microsoft Azure overview, Azure Data Lake Store, Azure Analysis Services, HDInsight, The ThingWorx overview, ThingWorx Core, ThingWorx Connection Services, ThingWorx Edge, ThingWorx concepts, Hadoop, Amazon S3, Apache Spark for data processing, Lambda Architectures, Handling change, Exploring and visualizing data
Unit 4
The Tableau overview, Techniques to understand data quality, Basic time series analysis, Get to know categories in the data, Solving industry-specific analysis problems, Decorating Your Data - Adding External Datasets to Innovate, Adding internal datasets, Adding external datasets, External datasets - geography, External datasets - demographic, External datasets - economic, Communicating with Others - Visualization and Dashboarding, Common mistakes when designing visuals, The Hierarchy of Questions method, Designing visual analysis for IoT data, Creating a dashboard with Tableau, Creating and visualizing alerts, Applying Geospatial Analytics to IoT Data, Why do you need geospatial analytics for IoT?, The basics of geospatial analysis, Vector-based methods, Raster-based methods, Storing geospatial data, Processing geospatial data, Forecasting, Edge computing, Fogg Computing
Unit 5
Security for IoT , Common challenges in OT security, formal risk analysis structures: OTAVE and FAIR, Hands on - Open source IoT Platforms: - Zetta, Kaa, Node-RED, Thinger
Reference Books: 1. Analytics for the Internet of Things (IoT), Andrew Minteer, 2017, Packt Publishing, ISBN: 9781787120730 2. https://www.zettajs.org/ 3. https://nodered.org/users/go-iot/

4. https://www.kaaproject.org/
5. https://thinger.io/
6. https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-overview.pdf
7. https://www.cisco.com/c/dam/en_us/solutions/trends/iot/docs/computing-solutions.pdf
8. https://www.cisco.com/c/en_in/solutions/computing/what-is-edge-computing.html
9. https://www.ibm.com/downloads/cas/0WOR6ORJ
10. https://learn.adafruit.com/alltheiot-protocols/coap
11. https://xiaozhon.github.io/course_tutorials/Coap_tutorial_RPi.pdf
12. https://raspberry-valley.azurewebsites.net/CoAP-Getting-Started/