M.SC. (DATA SCIENCE) COURSE CURRICULUM

(W.E.F. THE ACADEMIC YEAR 2026-2027)

DEPARTMENT OF COMPUTER SCIENCE

FACULTY OF SCIENCE

TELANGANA UNIVERSITY

NIZAMABAD

Approved in the meeting of the Board of Studies (BoS) in Computer Science, Telangana University, Nizamabad

TELANGANA UNIVERSITY, NIZAMABAD
FACULTY OF SCIENCE
M.SC. (DATA SCIENCE) COURSE CURRICULUM
POST GRADUATE PROGRAMME (20+22+24+22=88 CREDITS)

SEMESTER – I (20 CREDITS)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Data Structures	4	4	40	60	100
2	DSC	Operating Systems	4	4	40	60	100
3	DSC	Mathematical Foundations for Data Science	4	4	40	60	100
4	PR-1	Data Structures Lab	2	4	20	30	50
5	PR-2	Operating Systems lab	2	4	20	30	50
6	VAC	DevOps	2	2	-	50	50
7	IDC	Business Intelligence	2	2	20	30	50
		Total	**20**	**24**	**160**	**340**	**500**

DSC: Discipline Specific Course
CA: Continuous Assessment
ESE: End Semester Examination
VAC: Value-Added Course
IDC: Inter Disciplinary Course
PR: Practical

SEMESTER – II (22 Credits)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Data Mining	4	4	40	60	100
2	DSC	Advanced Artificial Intelligence	4	4	40	60	100
3	DSC	Statistical Inference	4	4	40	60	100
4	DSC-PR-1	Data Mining lab	2	4	20	30	50
5	DSC-PR-2	Advanced Artificial Intelligence Lab	2	4	20	30	50
6	SEC	Data Security and Privacy *(Department Specified)*	4	4	40	60	100
7	VAC	Quantum Computing Techniques	2	2	20	30	50
		Total	**22**	**26**	**200**	**350**	**550**

DSC: Discipline Specific Course
DSE: Discipline Specific Elective
PR: Practical
VAC: Value Added Course
SEC: Skill Enhancement Course
QT: Quantum Techniques

SEMESTER – III (24 CREDITS)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Optimization Techniques	4	4	40	60	100
2	DSC	Research Methodology	4	4	40	60	100
3	DSE	1. Web Mining 2. Deep Learning	4	4	40	60	100
4	DSC-PR-1	Optimization Techniques Lab	2	4	20	30	50
5	DSC-PR-2	1. Web Mining lab 2. Deep Learning lab	2	4	20	30	50
6	OEC	Open Elective (Select one interdisciplinary course).	4	4	40	60	100
7	VC	Internship/ MAD Apprenticeship/ Online / MOOCS/VC	2	2	-	50	50
8	ECA	ECA/ Mentorship /Tutorials	2	2	-	50	50
		Total	**24**	**28**	**180**	**420**	**600**

DSC: Discipline Specific Course
DSE: Discipline Specific Elective
RM: Research Methodology
PR: Practical
VC: Vocational Course
ECA: Extra Curricular Activities

*__OEC: Open Elective Course__- Students should select a course offered by other departments/ disciplines.

SEMESTER – IV (22 CREDITS)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Natural Languages Processing	4	4	40	60	100
2	DSC	IoT And Predictive Analytics	4	4	40	60	100
3	DSE	1. Cloud Computing 2. Big Data	4	4	40	60	100
4	DSC-PR-1	Natural Languages Processing Lab	2	4	20	30	50
5	DSE-PR-2	1. Cloud Computing Lab 2. Big Data Lab	2	4	20	30	50
6	RP	Research Project (Report (100) + Viva (50)	**6**	6	-	150	150
	For Research Project, a group of six students will be guided for 2 HPW, so the total workload will be the multiple of number of groups and 2 HPW						
Total			**22**	**26**	**140**	**410**	**550**

DSC: Discipline Specific Course
DSE: Discipline Specific Elective
RP: Research Project
PR=Practical

TELANGANA UNIVERSITY,
NIZAMABAD
FACULTY OF SCIENCE

M.SC. (DATA SCIENCE) COURSE CURRICULUM

SEMESTER – I (20 CREDITS)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Data Structures	4	4	40	60	100
2	DSC	Operating Systems	4	4	40	60	100
3	DSC	Mathematical Foundations for Data Science	4	4	40	60	100
4	PR-1	Data Structures Lab	2	4	20	30	50
5	PR-2	Operating Systems Lab	2	4	20	30	50
6	VAC	DevOps	2	2	-	50	50
7	IDC	Business Intelligence	2	2	20	30	50
		Total	**20**	**24**	**160**	**340**	**500**

DSC: Discipline Specific Course
AI: Artificial Intelligence
ESE: End Semester Examination
PR: Practical
VAC: Value-Added Course
CA: Continuous Assessment
IDC: Inter Disciplinary Course

MSCDS111		Data Structure		DS
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- Understand the basic concepts of data structures and algorithm analysis.
- Learn different searching, sorting, stack, and queue techniques.
- Understand linked lists, trees, and graph data structures and their operations.
- Develop problem-solving skills using appropriate data structures in java.

Course Outcomes

- Get idea about the concepts of data structures, algorithms, and complexity analysis.
- Apply searching, sorting, stack, and queue operations to solve problems.
- Implement linked lists, trees, and graph traversals using java.
- Select suitable data structures for efficient data organization and processing

UNIT -I

Fundamental Concepts: Introduction to Data Structures, Types of Data Structures, Implementation of data structures, Analysis of Algorithms, Complexity of algorithms: Space complexity, Time complexity Definition. Asymptotic Notation. Linear Data Structure using Arrays: Sequential Organization, Linear Data Structure Using Sequential Organization: Array ADT Addition, Multiplication, Sparse Matrix: Representation, Addition, Transpose. Searching: Search Techniques: Sequential search, Binary search. Sorting: Types of sorting, General sort concepts, Bubble sort, Insertion sort, Selection sort, Quick sort, Merge sort.

UNIT – II

Linked Lists: Introduction, Linked List: Comparison of sequential and Linked Organizations, Primitive operations, Realization of Linked Lists using arrays and dynamic memory management, Dynamic memory management, Doubly Linked List: Creation, Deletion, Insertion, Traversal. Stacks: Introduction to stack, Stack Abstract Data Type, operations using Stacks, Stack applications. Queues: Concept of Queues, Queue as Abstract Data Type, Circular Queue, Queue operations, Queue applications.

UNIT -III

Trees: Introduction, Basic terminology, General trees, Representation of a general tree, Types of Trees, Binary Tree, Properties, Binary Tree Abstract Data Type, Binary Tree Traversal Binary Search Tree, AVL Trees. Graphs: Introduction, Graph Abstract Data Type, Graph Representation, Graph traversals.

UNIT - IV

Design and Analysis of Algorithms: Introduction to algorithm design techniques, Performance analysis of algorithms, Recurrence relations, Divide and Conquer strategy, Binary Search, Merge Sort, Quick Sort, Greedy Method, Minimum Spanning Trees,

Shortest Path Algorithms, Dynamic Programming, Principles of Optimality, 0/1 Knapsack Problem, Longest Common Subsequence, Backtracking, N-Queens Problem, Graph Coloring, Branch and Bound Technique, Travelling Salesman Problem, Introduction to NP-Hard and NP-Complete Problems.

Text Books:

1. Fundamentals of Computer Algorithms, Ellis Horowitz, Sartaj Sahni and Rajasekharan, Universities press
2. Data Structures & Algorithms in Java by Robert Lafore, Sams Publishing.

Reference Books

1. Reema Thareja, Data structures using C, 2nd Edition, Oxford Publication.
2. D. S. Malik, *Data Structures Using C++*, Cengage Learning.
3. Adam Drozdek, *Data Structures and Algorithms in C++*, 4th Edition, Cengage Learning.

MSCDS112		Operating Systems		OS
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- Understand the basic concepts, structure, and services of operating systems.
- Learn process management, scheduling, multithreading, and interprocess communication.
- Understand synchronization, deadlocks, and memory management techniques.
- Gain knowledge of file systems, virtual memory, and disk management concepts.

Course Outcomes

- Get an idea on functions and structure of modern operating systems.
- Apply CPU scheduling, synchronization, and deadlock handling techniques.
- Analyze memory management and virtual memory methods in operating systems.
- Understand file system organization, storage management, and disk scheduling techniques.

UNIT - I

Introduction: Introduction to Operating Systems, Operating System Structure, Distributed Systems, Real Time OS, Batch Processing vs Multiprogramming, System Structures: Operating System Services, System Boot, System Interface, System Calls, Kernel, Process Management: Process concepts, Process State, Process Control Block (PCB), Cooperating Processes, and Inter Process Communication: Shared Memory, Message Passing, and Multithreaded Programming.

UNIT - II

CPU Scheduling: Scheduling Criteria, Scheduling Algorithms: FCFS, SJF, Priority, Round Robin, Multiple Processor Scheduling, Process Synchronization: Principle of Concurrency, Producer / Consumer Problem, Mutual Exclusion, Critical Section Problem, Semaphores, Classical Problem in Concurrency- Bounded Buffer Problem, Readers-Writers Problem, Dining Philosopher Problem, Deadlock: System model, Deadlock characterization, Prevention, Avoidance and detection, Recovery from deadlock.

UNIT -III

Memory Management: Swapping, Multiprogramming with fixed partitions, dynamic partitions, Paging, Segmentation, Virtual Memory Concepts, Demand Paging, Page fault, Page Replacement Algorithms and Limitations: FIFO (First In First Out), OPT (Optimal Page Replacement Algorithm), LRU (Least Recently Used), LFU (Least Frequently Used), MFU (Most Frequently Used), NRU (Not Recently Used). Thrashing,

UNIT - IV

Storage Management: File System: File attributes, Access Method: Sequential Access, Direct (Random) Access, Indexed Access, Allocation Methods: Contiguous Allocation,

Linked Allocation, and Indexed Allocation. Disk structure and disk performance parameters, Disk scheduling algorithms – FCFS (First Come First Served), SSTF (Shortest Seek Time First), SCAN, C-SCAN. Introduction to RAID, Need and advantages of RAID, RAID levels. I/O Management: Introduction to Input/Output Systems, Characteristics of I/O devices, Device controllers and device drivers, Interrupt-driven I/O, Direct Memory Access (DMA), I/O Management and Disk Scheduling: I/O devices, and I/O subsystems, I/O buffering, Disk storage and disk scheduling, RAID.

Text Books

1. Operating System Concepts by Abraham Silberschatz, Peter B. Galvin, Greg Gagne, Wiley Publications, 10th edition, 2018.

Reference Books

2. 1.Modern Operating Systems" by Andrew S. Tanenbaum and Herbert Bos, 2nd edition
3. 2.Operating Systems by Achyut S. Godbole, Tata McGraw Hill, 3rd edition, 2010.
4. 3.Operating Systems: Internals and Design Principles by William Stallings, Pearson Education, 9th edition, 2011.
5. 4.Operating Systems: A Concept-based Approach by Dhamdhere, D.M.,McGraw Hill, 2nd edition 6.

MSCDS113		Mathematical Foundations for Data Science		MFDS
DSC	**Credits: 4**	**HPW: 4**	**CA: 40**	**ESE: 60**

Course Objectives

- To provide fundamental knowledge of sets, relations, mathematical logic, and combinatorics required for data science applications.
- To develop analytical and problem-solving skills using recurrence relations, graph theory, and Boolean algebra concepts.
- To introduce mathematical structures and algorithms that support data representation, optimization, and computational thinking in data science.
- To familiarize students with graph-based models, logical reasoning, and Boolean functions used in computer science and artificial intelligence.

Course Outcomes

After completion of this course, the student will be able to:

- Apply concepts of sets, relations, equivalence relations, and mathematical logic in problem solving.
- Solve counting and combinatorial problems using permutations, combinations, binomial coefficients, and inclusion–exclusion principles.
- Analyze and solve recurrence relations and apply generating functions in mathematical and computational contexts.
- Interpret graph theoretic concepts such as trees, spanning trees, graph coloring, and traversal algorithms for practical applications.

UNIT-I

Sets and Relations: Set Operations, Representation and Properties of Relations, Equivalence Relations, Partially Ordering.

Mathematical Logic: Fundamentals of logic, Logical inferences, Methods of proof of an implication, first order logic and other methods.

UNIT-II

Combinatorics: Basic of counting, Permutations and Combinations, Enumeration of Permutations and Combinations with Repetitions and Constrained Repetitions, Binomial coefficients, Principle of Inclusion-Exclusion.

Recurrence Relations: Introduction, solving linear recurrence relations, generating functions.

UNIT-III

Graph Theory: Basic concepts of Graphs, Isomorphic graphs and simple problems, Trees and their properties, spanning tree, breadth and depth first search, minimal spanning tree constructions, Kruskal’s algorithm, Planar graphs, Euler’s formula, Hamiltonian graphs, simple problems and applications, Graph coloring, Chromatic numbers, four color problem.

UNIT – IV

Boolean algebra: Introduction to Boolean algebra functions, Switching Mechanisms, Minimization of Boolean functions. Applications to Digital Computer design.

Textbooks:

1. Discrete Mathematics for Computer Scientists and Mathematicians. Joe L. Mott, Abraham Kandel, Theoder P. Baker, Prentic-Hall of India Private Limited.
2. Discrete Mathematics by Bernord kolnan, Robert C. Busby and Sharon Ross.
3. Fundamentals of Mathematical Statistics. S.C. Gupta, V.K. Kapoor. Sultan Chand & Sons. New Delhi.

MSCDS114		Data Structures lab		DS-PR1
DSC	Credits: 2	HPW: 4	CA: 20	ESE: 30

List of Experiments:

1. Write a program to implement the stack operations using an array
2. Write a program to implement the Queue operations using an array
3. Write a program to implement the stack operations using a singly linked list.
4. Write a program for linked list implementation of Queue.
5. Write a program for postfix expression evaluation.
6. Write a program to convert prefix to postfix.
7. Write a Program to Implement Binary Search Tree operations (insertion, deletion, search, count nodes, check empty)
8. Write a program to implement deque using double linked list.
9. Write a program to implement Linear Search
10. Write a program to implement Binary Search
11. Write a program for merge sort.
12. Write a program for Quick sort.
13. Implement Prim's Algorithm to find the Minimum Spanning Tree of a weighted graph.
14. Implement Kruskal's Algorithm to find the Minimum Spanning Tree of a weighted graph.
15. Implement Dijkstra's Algorithm to find the shortest path from a source vertex to all other vertices.
16. Implement the 0/1 Knapsack Problem using Dynamic Programming.
17. Implement the Fractional Knapsack Problem using the Greedy Method.
18. Implement the N-Queens Problem using Backtracking.

Note: Students can execute the programs using C/python/ Java

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

MSCDS115		Operating System lab		OS-PR2
DSC	Credits: 2	HPW: 4	CA: 20	ESE: 30

List of Experiments:

1. Write a program to execute basic Linux commands using system calls (ls, pwd, date).
2. Write a C program to create a child process using fork() and display process IDs of parent and child.
3. Write a program to demonstrate the use of exec() family of system calls.
4. Write a program to demonstrate process synchronization using wait() and waitpid().
5. Write a program to implement Inter Process Communication using pipes.
6. Write a program to implement Inter Process Communication using shared memory.
7. Write a program to create and manage multiple threads using Pthreads.
8. Write a program to simulate First Come First Serve (FCFS) CPU scheduling algorithm
9. Write a program to simulate Shortest Job First (SJF) scheduling algorithm.
10. Write a program to simulate Round Robin (RR) scheduling algorithm.
11. Write a program to solve the Producer-Consumer problem using semaphores.
12. Write a program to solve the Readers-Writers problem.
13. Write a program to solve the Dining Philosophers problem.
14. Write a program to demonstrate Peterson's solution for mutual exclusion.
15. Write a program to implement Banker's Algorithm for deadlock avoidance.
16. Write a program to detect deadlock in a system.
17. Write a program to simulate FIFO Page Replacement Algorithm.
18. Write a program to simulate LRU Page Replacement Algorithm.
19. Write a program to simulate Optimal Page Replacement Algorithm.
20. Write a program to demonstrate Paging technique (address translation).
21. Write a program to demonstrate Segmentation.
22. Write a program to simulate Demand Paging and Page Fault handling.
23. Write a program to illustrate Thrashing.
24. Write a program to perform file operations using system calls (open, read, write, close).
25. Write a program to demonstrate directory operations (create, remove, list).
26. Write a program to simulate file allocation methods (sequential, indexed, linked).
27. Write a program to implement free space management using bit vector.
28. Write a program to simulate FCFS Disk Scheduling Algorithm.
29. Write a program to simulate SSTF Disk Scheduling Algorithm.
30. Write a program to simulate SCAN Disk Scheduling Algorithm.
31. Write a program to simulate C-SCAN Disk Scheduling Algorithm.
32. Write a program to demonstrate RAID concept (simulation).

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

MSCDS116		DevOps		DO
VAC	Credits: 2	HPW: 2	CA: --	ESE: 50

Course Objectives

- To understand the foundations, culture, and principles of DevOps.
- To study collaboration practices and team dynamics in DevOps environments.
- To analyze affinity, tooling, and automation strategies for effective software delivery.
- To understand scaling challenges and organizational transformation through DevOps practices.

Course Outcomes

- Understand the concepts of Agile, DevOps, ITIL, Continuous Delivery, Scrum, and Kanban in modern software development environments.
- Analyze DevOps lifecycle practices, continuous testing methods, software architecture patterns, and microservices for business agility and resilient systems.
- Apply DevOps tools and automation techniques to implement continuous integration, delivery pipelines, release management, and system updates effectively.
- Design and recommend scalable DevOps solutions for enterprise environments while addressing governance, security, and continuous improvement requirements.

UNIT-I

Foundations of DevOps: The Big Picture: What Is DevOps? The DevOps Equation, The Advent of Software engineering, Proprietary Software and Standardization, The Age of the Network, The Growth of Software Development Methodologies, Agile Infrastructure.
Software development models and DevOps: DevOps Lifecycle for Business Agility, DevOps, and Continuous Testing. DevOps influence on Architecture: Introducing software architecture, The monolithic scenario, Architecture rules of thumb, The separation of concerns, Handling database migrations, Microservices, and the data tier, DevOps, architecture, and resilience. Foundational Terminology and Concepts: Software Development Methodologies, Waterfall, Agile, Scrum, ITIL, COBIT, Lean, Version Control, Test-Driven Development, Continuous Integration, Continuous Delivery, Continuous Deployment, Configuration Management, Cloud Computing, Infrastructure Automation, Containers, Retrospectives, Postmortems, Blamelessness, Organizational Learning.

UNIT-II

Affinity and Tools: Affinity: From Individuals to Teams, People Networks, Team Culture, Team Cohesion, Diversity and Inclusion, Organizational Structure, Building Team Empathy, Improving Team Communication, Benefits of Improved Affinity, Trust and Innovation. Tools: Ecosystem Overview, Software Development Tools, Local

Development Environments, Version Control, Artifact Management, Infrastructure Automation, Test and Build Automation, Monitoring, Metrics, Logging and Alerting.
Scaling and Organizational Transformation: Scaling: Inflection Points, Enterprise DevOps, Organizational Growth, Team Flexibility, Organizational Lifecycle, Release Cycles, Team Scaling, Knowledge Sharing, Scaling for Organizations. Scaling: Misconceptions and Troubleshooting, Executive Buy-In, Resource Constraints, Team Management Challenges, Organizational Change. Building Bridges with the Four Pillars of Effective DevOps, Values and Culture, Engineering Exchanges, Human Connections, Healthy and Unhealthy Systems.

Textbook

1. Effective DevOps: Building a Culture of Collaboration, Affinity, and Tooling at Scale by Jennifer Davis and Ryn Daniels.

Reference Books

1. Gene Kim, Jez Humble, Patrick Debois, and John Willis, *The DevOps Handbook: How to Create World-Class Agility, Reliability, and Security in Technology Organizations*, IT Revolution Press.
2. Jez Humble and David Farley, *Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation*, Addison-Wesley Professional.
3. Nicole Forsgren, Jez Humble, and Gene Kim, *Accelerate: The Science of Lean Software and DevOps – Building and Scaling High Performing Technology Organizations*, IT Revolution Press.

MSCDS117		Business Intelligence		BI
IDC	Credits: 2	HPW: 2	CA: 20	ESE: 30

Course Objectives

- To introduce the fundamentals of Business Intelligence and Data Warehousing concepts.
- To understand OLTP, OLAP architectures, and multidimensional data models.
- To develop knowledge of ETL, data integration, and data quality techniques.
- To provide understanding of enterprise reporting, dashboards, and balanced scorecards.

Course Outcomes

After successful completion of the course, students will be able to:

- Explain Business Intelligence concepts, OLTP/OLAP systems, and data warehouse architecture.
- Design multidimensional data models using star and snowflake schemas.
- Apply ETL, data integration, and data quality techniques in data warehousing environments.
- Develop and interpret enterprise reports, dashboards, and balanced scorecards for decision-making.

UNIT – I

Introduction to Business Intelligence and Data Warehousing: Introduction to Business Intelligence (BI), Digital data and its types – structured, semi-structured and unstructured data. Introduction to Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP), Different OLAP architectures: MOLAP, ROLAP, HOLAP, Comparison of OLTP and OLAP. BI definitions, need and importance of BI, BI Component Framework, Business Applications of BI. Data Warehouse: Definition, reasons and goals of a data warehouse, Components and characteristics of a data warehouse. Introduction to Extraction-Transformation-Loading (ETL), Data Integration – needs, advantages and common approaches. Introduction to Data Quality and Data Profiling concepts and applications.

UNIT – II

Data Modeling and Enterprise Reporting**:** Introduction to Data and Dimension Modeling, Multi-dimensional Data Model, ER Modeling vs. Multi-dimensional Modeling. Concepts of dimensions, facts, cubes, attributes and hierarchies. Typical dimensional models – Star Schema and Snowflake Schema. Basics of Enterprise Reporting: Features of good reporting, Common report layout types, Report delivery formats, Report standardization and presentation practices. Introduction to Balanced Scorecard and Enterprise Dashboards, Balanced Scorecards vs. Enterprise Dashboards.

Text Books:

1. R.N. Prasad, Seema Acharya, Fundamentals of Business Analytics, Wiley India Ltd.

Reference Books:

1. Mike Biere, Business Intelligence for the Enterprise, Prentice Hall Professional
2. David Taniar, Progressive methods in Data Warehousing and Business Intelligence: Concepts and competitive analytics, Idea Group Inc.

TELANGANA UNIVERSITY, NIZAMABAD
FACULTY OF SCIENCE

M.SC. (DATA SCIENCE) COURSE CURRICULUM

SEMESTER – II (22 CREDITS)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Data Mining	4	4	40	60	100
2	DSC	Advanced Artificial Intelligence	4	4	40	60	100
3	DSC	Statistical Inference	4	4	40	60	100
4	DSC-PR-1	Data Mining lab	2	4	20	30	50
5	DSC-PR-2	Advanced Artificial Intelligence Lab	2	4	20	40	50
6	SEC	Data Security and Privacy *(Department Specified)*	4	4	40	60	100
7	VAC	Quantum Computing Techniques	2	2	20	30	50
		Total	**22**	**26**	**200**	**350**	**550**

DSC: Discipline Specific Course
DSE: Discipline Specific Elective
PR: Practical
VAC: Value Added Course
SEC: Skill Enhancement Course
QT: Quantum Techniques

MSCDS121		Data Mining		DM
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- Study data warehouse principles and its working
- Learn Data mining concepts and understand Association Rule Mining
- Study Classification Algorithms
- Gain knowledge of how data is grouped using clustering techniques.

Course Outcomes

- Comparison of functional differences between data warehouse and database systems.
- Ability to perform the pre-processing of data and apply mining techniques on it.
- Capability to identify the association rules, classification and clusters in large data sets.
- Skills to solve real world problems in business and scientific information using data mining.

UNIT - I

Introduction: Why Data Mining? What is Data Mining? What kinds of data can be mined? What kinds of patterns can be mined? Which technologies are used? Which kinds of applications are Targeted? Major issues in Data Mining. Getting to know your data: Data objects and attributed types. Basic statistical descriptions of data. Measuring data similarity and dissimilarity.

UNIT-II

Data Pre-Processing: Data Cleaning, Data Integration - Data reduction: Overview, Attribute subset selection, clustering, sampling, Data cube Aggregation, Histograms. Data Transformation and Data Discretization and concept Hierarchy Generation.

Data Warehouse: Basic Concepts, Data Warehouse modeling, Data Warehouse Design. Data Warehouse implementation- Data cube implementation overview and OLAP server architecture, Attribute oriented induction.

UNIT-III

Classification: Basic concepts, Decision tree induction, Bayes classification methods, Classification: Advance methods, Bayesian Belief Network, Classification by back propagation, Support vector machine,

UNIT-IV

Cluster Analysis: Concepts and Methods: Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Evaluation of clustering. Mining Complex Data Types, Other Methodologies of Data Mining, Data Mining Applications, Data Mining and Society.

Text books:

1. Jiawei Han , Micheline Kamber, Jin Pei, Data Mining: Concepts & Techniques, 3rd Edition., Morgon Koffman ,2011
2. Vikram Pudi P.Radha Krishna, Data Mining, Oxford University Press, 1st Edition, 2009.
3. 3. Pang-Ning Tan, Michael Steinbach, Vipin kumar, Introduction to Data Mining, Pearson Education, 2008.

MSCDS122		Advanced Artificial Intelligence		AAI
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To understand the foundations of Artificial Intelligence, intelligent agents, and problem-solving techniques.
- To learn search algorithms, heuristic methods, knowledge representation, and reasoning techniques in AI.
- To develop skills in machine learning, including regression, classification, clustering, and model evaluation.
- To apply modern AI techniques, foundation models, and prompt engineering for intelligent decision-making and data-driven applications.

Course Outcomes

After successful completion of the course, students will be able to:

- Apply search algorithms, heuristic techniques, and knowledge representation methods to solve AI problems.
- Design intelligent systems using logical reasoning, game-playing strategies, and machine learning approaches.
- Build and evaluate regression, classification, clustering, and ensemble learning models using appropriate techniques.
- Utilize foundation models, language model evaluation metrics, and prompt engineering methods for real-world AI applications.

Unit - I

Introduction: Definitions of Artificial Intelligence, The Foundations of Artificial Intelligence, The History of Artificial Intelligence, The State of the Art. problem solving in AI, Defining an AI Problem as state space search, production systems, AI problem characteristics. Search techniques: Breadth first search, Depth first search.

Unit– II

Heuristic Search Techniques: Generate and test, hill climbing, best first search, Heuristic function applications, problem Reduction, simulated annealing. Knowledge Representation in AI: knowledge representations approaches & issues in knowledge representation, Knowledge Representation using predicate logic, forward and backward reasoning, semantic nets, frames, scripts & conceptual Dependency.

Unit – III

Game Playing: The Minimax Search procedure, Adding Alpha-Beta Cutoffs, Additional Refinements, Iterative Deepening. Reasoning in Uncertain Situations: Introduction to Non monotonic Reasoning, Logic-Based Abductive Inference. Introduction to Machine Learning: Introduction, Classic and Adaptive machines, Learning Types-Supervised, Unsupervised, deep learning, bio-inspired adaptive systems, Machine Learning, and big data.

Unit – IV

Supervised Learning - Introduction to supervised learning, Regression: Introduction of regression, Regression algorithms: Simple linear regression, Multiple linear regression, Logistic regression, Maximum likelihood estimation.

Classification: Classification model and learning steps, Classification algorithms: Naïve Bayes classifier, Distance measures, Support vector machines, Random Forest. ROC curve. Naïve Bayes and Discriminant Analysis: Bayes theorem, Ensemble Learning: Introduction, to Ensemble Learning-Random Forests, AdaBoost, Gradient Tree Boosting, Support Vector Machines: Linear SVM, Kernel based Classification. Clustering Fundamentals: Basics, K-means, Evaluation methods, Spectral Clustering, Hierarchical Clustering. Evaluation Methodology: Challenges of Evaluation Foundation Models, understanding languages modeling metrics. Prompt engineering: Introduction to prompting, In-context learning: Zero-shot and Few-Shot, System Prompt and User Prompt.

Text Books

1. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th Edition, Pearson Education.
2. Tom M. Mitchell, *Machine Learning*, McGraw-Hill Education.
3. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd Edition, O'Reilly Media.

Reference Books

1. Subramanian Chandramouli, Amit Kumar Das, and Saikant Dutt, *Machine Learning*, Pearson Education, 2018.
2. Chip Huyen, *AI Engineering: Building Applications with Foundation Models*, O'Reilly Media, 2025.
3. Ethem Alpaydin, *Introduction to Machine Learning*, 4th Edition, MIT Press.

MSCDS123		Statistical Inference		SI
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To understand the fundamental concepts of statistical inference and estimation methods.
- To learn point estimation, interval estimation, and hypothesis testing techniques.
- To develop knowledge of MP, UMP, and parametric testing procedures.
- To study non-parametric statistical methods and their practical applications.

Course Outcomes

After successful completion of the course, students will be able to:

- Apply estimation methods and construct confidence intervals for statistical parameters.
- Perform hypothesis testing using parametric and non-parametric techniques.
- Analyze statistical data using MP, UMP, and goodness-of-fit tests.
- Interpret and apply non-parametric inference methods in research and real-world problems.

UNIT-I

C.R. Rao's Definition of Statistical Inference.

Estimation Theory**:** Criteria of a good estimator: Unbiasedness, consistency, efficiency and sufficiency with examples. Cramer-Rao inequality, Rao-Blackwell theorem, Fisher Information, and Lehmann-Scheffe theorem. Simple Problems on UMVUE.

UNIT-II

Methods of Estimation: Method of maximum likelihood estimation and its Properties, Simple problems on MLE. Jackknife, Bootstrap resampling methods. Estimation of bias and standard deviation of point estimation by the Jackknife & bootstrap methods with examples, U-Statistic, Kernal and examples. Interval estimation, confidence level CI using pivots and shortest length CI and example problems.

UNIT-III

Testing of Hypotheses: Fundamental notions of hypothesis testing – Statistical hypothesis, Critical region, types of errors, Level of significance, test function, randomized and non-randomized tests, power function, most powerful test, Neyman – Pearson fundamental lemma, MLR families and uniformly most powerful tests for one parameter exponential families. Parametric tests: Test of significance based on means, variances and goodness of fit using t-test, F-test and Chi-square test.

UNIT-IV

Non parametric tests: Definition, assumptions, advantages and disadvantages of non-parametric inference. Non-parametric methods for one and two Sample problems based on sign test, Wilcoxon signed Rank test, Wilcoxon Mann-Whitney U test, Kolmogorov-

Smirnov test, Kruskal Wallis test and Wald – Wolfowitz Run test. Non-parametric Density Estimation**:** Concepts of Non-parametric estimation, Density estimates, Survey of existing methods. Rosenblatt's naïve density estimator, its bias and variance. Consistency of Kernel density estimators and its MSE.

Text Books:

1. Rohatgi, V.K.: *An Introduction to Probability Theory and Mathematical Statistics*, Wiley.
2. Gibbons: *Non-Parametric Statistical Inference*, TMH.
3. Lehmann, E. L.: *Testing of Hypothesis*, John Wiley.
4. Goon, Gupta and Das Gupta: *Outlines of Statistics, Vol. II*, World Press.
5. C.R. Rao: *Linear Statistical Inference*, John Wiley.
6. V.K. Kapoor and S.C. Gupta: *Fundamentals of Mathematical Statistics*, Sultan Chand & Sons.
7. Parimal Mukhopadhyay: *Mathematical Statistics*, New Central Book Agency.

MSCDS124		Data Mining lab		DM-PR1
DSC	Credits: 2	HPW: 4	CA: 20	ESE: 30

(Using Python, Java, WEKA or any opensource data mining tool)

List of Experiments:

1. List all the categorical (or nominal) attributes and the real-valued attributes separately.
2. One type of model that you can create is a Decision Tree -train a Decision Tree using the complete dataset as the training data. Report the model obtained after training.
3. Implement classification with a user define data set.
4. Describe what is cross -validation briefly. Train a Decision Tree again using cross validation and report your results.
5. Implement decision tree with cross validation.
6. Write a program to Generate Association rules by using Apriori algorithm
7. Write a program to implement Naïve Bayesian classification
8. Write a program to implement K-Means clustering algorithm
9. Write a program to implement K-Mediods clustering algorithm
10. Write a program to implement DBSCAN algorithm.

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

MSCDS125		Advanced Artificial Intelligence Lab		AAI-PR2
DSC	Credits: 2	HPW: 4	CA: 20	ESE: 30

List of Experiments

1. Implement BFS and DFS both algorithms, compare traversal order, path lengths, and time taken. Visualise the explored nodes.

2. Apply hill climbing and simulated annealing to an optimisation problem (e.g., N-queens or TSP). Compare how simulated annealing escapes local optima that trap hill climbing.

3. Implement A* using Manhattan and Euclidean heuristics for pathfinding; compare performance against BFS on the same problem.

4. Demonstrate forward chaining (data-driven) and backward chaining (goal-driven) reasoning using propositional / predicate logic rules.

5. Build a semantic network and a frame-based knowledge base for a small domain (e.g., animals or vehicles). Perform simple inheritance and slot-filling queries.

6. Implement a two-player game (Tic-Tac-Toe or Connect-4) AI using the minimax algorithm. Add alpha-beta cutoffs and measure the reduction in nodes evaluated.

7. Demonstrate supervised learning (classification) and unsupervised learning (clustering) on small datasets. *Reinforcement learning demo removed* — not in syllabus.

8. Perform EDA on a dataset (distributions, correlations, outliers). Apply preprocessing: handle missing values, encode categorical features, normalise data.

9. Implement simple linear, multiple linear, polynomial, and logistic regression on appropriate datasets. Interpret coefficients and compare fit. (Merges original Exps 10 & 11.)
10. Implement Naïve Bayes (Bayes theorem), KNN (distance measures), and Decision Tree classifiers. Compare accuracy, precision, recall, and ROC curves across all three.

11. Implement Random Forest, AdaBoost, and Gradient Tree Boosting. Implement linear and kernel-based SVM. Visualise decision boundaries and compare all models.

12. Implement K-means, hierarchical, and spectral clustering on a dataset; evaluate using silhouette score. Then perform cross-validation and compare multiple classification models using standard metrics.

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

MSCDS126		Data Security and Privacy		DSP
SEC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To understand the fundamentals of cryptography and network security principles.
- To learn symmetric and asymmetric encryption techniques for secure communication.
- To study authentication, hashing, and digital signature mechanisms for data protection.
- To explore network and web security methods including firewalls, IP security, and secure transactions.

Course Outcomes

After successful completion of the course, students will be able to:

- Analyze cryptographic principles, security models, and network attacks.
- Apply encryption, key exchange, and authentication techniques for secure communication.
- Implement hashing, digital signatures, and cryptographic algorithms for data security.
- Evaluate security mechanisms used in email systems, web applications, and network infrastructures

UNIT - I

Introduction: Attacks, security Services, Mechanisms, A Model for Network Security Introduction to Number Theory: Divisibility and the Division Algorithm, the Euclidean Algorithm, Modular Arithmetic, Fermat's and Euler's Theorem, the Chinese Remainder Theorem, Discrete Logarithms. Symmetric Cipher: Classical Encryption Techniques, Symmetric Cipher Model, Substitution Techniques, Transposition Techniques, Steganography.

UNIT - II

Block Cipher and the Data Encryption: Block Cipher Principles, the Data Encryption Standard, Multiple Encryption and Triple DES, Stream Cipher-RC4, Block Cipher Modes of Operation, Advanced Encryption Standard. Asymmetric Cipher: Principles of Public-Key Cryptosystems, the RSA Algorithm, Diffie-Hellman Key Exchange, Elgamal cryptographic system, Elliptic Curve Cryptography, Traffic Confidentiality: link encryption, end to end encryption.

UNIT –III

Cryptographic Data Integrity Algorithms and Message Authentication: Message Authentication Requirements, Message Authentication Function, Message Authentication Codes, Hash Functions, simple Hash functions, Hash function properties, Secure Hash Algorithm, SHA-512, HMAC, Digital Signatures: Properties, Attacks and forgeries, Digital signature Requirements, Kerberos, Firewalls.

UNIT –IV

Electronic Mail Security: Email Components, email format, Email Threats and security, S/MIME. IP Security: IP Security Overview, IP Security Architecture, Encapsulating Security Payload. Web and System Security: malicious software, Intruders, Viruses, Secure Electronic Transaction, HTTPS.

Text Book:

1. Cryptography and Network Security principles and Practice 7th Edition by William Stallings (Pearson Asia).

Reference Books:

1. Davies & Price: Security for Computer Networks - Wiley
2. Mayer & Matyas: Cryptography - Wiley B. Schneider: Applied Cryptography - (John Wiley).

MSCDS127		Quantum Computing Techniques		QCT
VAC	Credits: 2	HPW: 2	CA: 20	ESE: 30

Course Objectives

- To introduce the principles of quantum computation and quantum information.
- To understand qubits, quantum gates, and quantum circuits.
- To study fundamental quantum algorithms and complexity concepts.
- To explore quantum cryptography and emerging applications of quantum computing.

Course Outcomes

After successful completion of the course, students will be able to:

- Explain the foundations of quantum computing and distinguish it from classical computing.
- Analyze quantum states, quantum gates, and quantum circuits.
- Understand the operation and significance of major quantum algorithms.
- Evaluate the applications and future prospects of quantum computing technologies.

UNIT – I

Fundamentals of Quantum Computing. Classical Information and Computation: Bits, Boolean Logic Gates, Reversible Computation, Reversible Logic Gates, Computational Complexity and Limitations of Classical Computing. Introduction to Quantum Computing: Historical Background, Quantum Information, Quantum States and Qubits, Bra-Ket Notation, Superposition Principle, Quantum Measurement. Quantum Gates and Circuits: Single-Qubit Gates (Pauli-X, Pauli-Y, Pauli-Z, Hadamard and Phase Gates), Multi-Qubit Gates (CNOT and Toffoli Gates), Quantum Gates as Matrices, Quantum Circuits and Circuit Representation.

UNIT – II

Quantum Algorithms and Applications. Multiple Qubit Systems: Product States, Entangled States, Quantum Entanglement and its Significance. Quantum Algorithms: Introduction to Quantum Algorithms, Deutsch-Jozsa Algorithm, Grover's Search Algorithm, Shor's Factoring Algorithm, Circuit Complexity and Query Complexity. Quantum Programming Fundamentals: Quantum Programming Concepts and Quantum Computing Platforms. Quantum Cryptography and Applications: Quantum Key Distribution, Introduction to Quantum Cryptography, Applications of Quantum Computing in Artificial Intelligence, Optimization, Cybersecurity and Scientific Computing.

Text Book

1. Thomas G. Wong, Introduction to Classical and Quantum Computing, Rooted Grove Publisher.
2. Quantum Computation and Quantum Information – Michael A. Nielsen & Isaac L. Chuang, Cambridge University Press.

Reference Books

1. Michael A. Nielsen and Isaac L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press.
2. Chris Bernhardt, Quantum Computing for Everyone, MIT Press.
3. N. David Mermin, Quantum Computer Science: An Introduction, Cambridge University Press.
4. Jack D. Hidary, Quantum Computing: An Applied Approach, Springer.

TELANGANA UNIVERSITY,
NIZAMABAD
FACULTY OF SCIENCE

M.SC. (DATA SCIENCE) COURSE CURRICULUM

SEMESTER – III (24 CREDITS)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Optimization Techniques	4	4	40	60	100
2	DSC	Research Methodology	4	4	40	60	100
3	DSE	1. Web Mining 2. Deep Learning	4	4	40	60	100
4	DSC-PR-1	Optimization Techniques Lab	2	4	20	40	50
5	DSE-PR-2	1. Web Mining lab 2. Deep Learning lab	2	4	20	40	50
6	OEC	Open Elective (students will select from other disciplines)	4	4	40	**60**	100
7	VC	Internship/ MAD Apprenticeship/ Online / MOOCS/VC	2	2	-	50	50
8	ECA	ECA/ Mentorship /Tutorials	2	2	-	50	50
		Total	**24**	**28**	**180**	**420**	**600**

DSC: Discipline Specific Course
DSE: Discipline Specific Elective
RM: Research Methodology
PR: Practical
VC: Vocational Course
ECA: Extra Curricular Activities

*__OEC: Open Elective Course__- Students should select a course offered by other departments/ disciplines.

MSCDS211		Optimization Techniques		OT
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To introduce the fundamental concepts and mathematical techniques of optimization and decision making.
- To develop understanding of linear programming, transportation, assignment, and sequencing models.
- To study queueing systems, game theory, and network optimization techniques.
- To train students in solving real-life resource allocation and decision-making problems using optimization methods.

Course Outcomes

After successful completion of the course, students will be able to:

- Apply linear programming and duality techniques to solve optimization problems.
- Analyze transportation, assignment, sequencing, and queueing models for decision making.
- Solve problems using game theory, dynamic programming, and network techniques.
- Use optimization methods for effective resource allocation and real-world applications.

UNIT – I

Linear Programming Problem: Formulation of linear programming problem, graphical solution method, Simplex method, Charnes Big-M method, two-phase simplex method. Duality: Concepts of primal and dual LPP, fundamental properties of duality, solution of LPP using its dual, dual simplex method with examples.

UNIT – II

Transportation problem: Concept of transportation problem, TPP as a special case of LPP, Initial basic feasible solutions by North-West Corner method, Matrix minimum method and Vogel's approximation method (VAM). Optimal solution of TPP through MODI method for balanced and unbalanced transportation problems. Assignment problem: Concept of Assignment problem, Assignment problem as special case of TP and LPP. Unbalanced assignment problem, optimal solution using Hungarian method.

UNIT - III

Sequencing Problems: Concept of Sequencing Problems, Johnson's method for optimal sequence of 'n' jobs on k machines where k = 2, 3,…,n without passing. Queueing theory: Description of queueing problem, characteristics of queueing process, steady state conditions of queueing, (M/M/1): (FIFO) queueing system, Poisson queues with exponential service times, traffic intensity ratio, average queue length, waiting time distribution, Average waiting time of an arrival, queueing system as Birth – Death process.

UNIT – IV

Game Theory: Concepts of a game, strategies, game as a special case of LPP, two-person zero-sum finite games, pay-off matrix, maximin-minimax principle, saddle point, pure and mixed strategies, games without saddle points and their solutions using graphical method, Dominance property techniques and algebraic method. Networks: Basic concepts, constraints in Networks, construction of Networks, Time calculation in Networks, PERT, CPM, Network Problems.

Text Books

1. Kanti Swarup, P. K. Gupta, and Man Mohan, *Operations Research*, Sultan Chand & Sons.
2. Hamdy A. Taha, *Operations Research: An Introduction*, Pearson Education.
3. R. Panneerselvam, *Operations Research*, Prentice Hall of India.

Reference Books

1. S. I. Gass, *Linear Programming: Methods and Applications*, McGraw-Hill.
2. Frederick S. Hillier and Gerald J. Lieberman, *Introduction to Operations Research*, McGraw-Hill Education.
3. Hira and Gupta, *Operations Research*, S. Chand Publishing.

MSCDS212		Research Methodology		RM
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To understand the fundamental concepts and importance of research methodology in computer science.
- To formulate research problems and design appropriate research methodologies.
- To develop skills in data collection, analysis, and interpretation using statistical and computational methods.
- To communicate research findings effectively through academic writing, presentations, and publications.

Course Outcomes

After successful completion of the course, students will be able to:

- Identify and apply appropriate research methodologies for computer science problems.
- Design and conduct research studies with ethical considerations and suitable data analysis techniques.
- Analyze and interpret research data using statistical tools and programming methods.
- Prepare research papers, presentations, and submissions for conferences and journals following academic standards.

Unit – I

Research Problem and Reporting Writing, Objectives of research, selecting the problem, Research design, Review of literature, Technical report writing, Research project proposal, Data interpretation, Gantt chart, Research project report, Manuscript writing, Plagiarism and Software tools, Acknowledgement, Citation styles. Bibliography.

Unit – II

Qualitative Research Methods Qualitative Data Collection and Analysis: Qualitative research methods, Data collection techniques: Interviews focus groups, Coding and interpretation of qualitative data. Content Analysis: What is content analysis? Types of content analysis (e.g., thematic analysis, textual analysis), Conducting content analysis in computer science research, Using software for content analysis. Mixed-Methods Research and Reporting: Combining quantitative and qualitative methods, Reporting research findings, Ethical considerations in mixed-methods research

Unit –III

Research Communication and Publishing Writing Research Papers: Structure of a research paper, Writing the abstract, introduction, literature review, and methodology sections, Citation and referencing. Research Presentations and Visualizations: Preparing effective research presentations, Visual aids and slide design, presenting quantitative and qualitative findings Publishing Research: Journals vs. conferences, the peer-review process, Preparing

manuscripts for submission, Ethical considerations in publishing including data privacy and security in publications.

Unit– IV

MS-Word and LaTeX: Manuscript preparation using MS-Word and LaTeX, Sample document, type style, environments, lists, centering, tables, verbatim, vertical and horizontal spacing, equation environments, fonts, hats and underlining, braces, arrays and matrices, customized commands, maths styles, document classes and overall structure, titles for documents, sectioning commands, packages, inputting files, pictures, making a bibliography, making an index, preparation of PPT slides

Text Book:

1. C.R. Kothari, Research Methodology Methods and Techniques, 2/e, Vishwa Prakashan, 2006.
2. The Unified Modeling Language User Guide by Grady Booch
3. Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development by Craig Larman

Reference Books

1. Bendat and Piersol, Random Data: Analysis and Measurement Procedures, Wiley Interscience, 2001.

MSCDS213		1. Web Mining		WM
DSE	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To provide fundamental knowledge of Web Data Mining and Information Retrieval systems.
- To understand web preprocessing, indexing, retrieval models, and web ranking techniques.
- To study web usage mining, user behavior analysis, and recommender systems.
- To explore applications of web mining in search engines, opinion mining, social networks, and business intelligence.

Course Outcomes

After successful completion of the course, students will be able to:

- Explain the concepts, architecture, and applications of Web Data Mining and Information Retrieval.
- Apply indexing, preprocessing, ranking, and web crawling techniques for efficient web search.
- Analyze web usage data using clustering, classification, and pattern mining methods.
- Design recommender systems and evaluate web mining applications such as spam detection, opinion mining, and social network analysis.

Unit - I

Web Data Mining – Introduction, Basic concepts of Information Retrieval, Information Retrieval Models-Boolean Model, Vector-Space Model, Statistical Language Model, Relevance Feedback. Web Page Pre-Processing, Indexing: Stopword Removal, Stemming, Web-page preprocessing Duplicate Detection, Inverted Index and its Compression, Latent Semantic Indexing;

Unit - II

Web Page Ranking and Search: Algorithms for ranking- PageRank Algorithm, HITS; Web Search- Using similarity scores, Using Rank positions; combining multiple rankings; Web Search Engines.

Web Crawling: Types of crawlers- Breadth first crawlers, Preferential crawlers, Universal crawlers, focused crawlers, topical crawlers; Implementation issues in crawling, Crawler ethics and conflicts

Unit - III

Web Usage Mining- Sources and types of data, Data modelling for web usage mining, discovery and analysis of web usage patterns, Discovery and Analysis of Web Usage Patterns- session and visitor analysis, Cluster analysis and Visitor Segmentation, Analysis of Sequential and Navigation patterns, Classification and prediction based on web user

transactions, Recommender Systems, Collaborative filtering- Filtering using K-NN, Filtering using Association Rules, Query Log Mining

Unit -IV

Applications of Web Mining: Web Spamming- Content Spamming, Link Spamming, Hiding techniques, Combating Spam; Social Network Analysis and Community Discovery; Mining Opinion from the Web; Business Applications of Web Mining

Textbooks:

1. Web Data Mining: Exploring Hyperlinks, Contents and Usage Data Bing Liu Springer, 2nd edition
2. Mining the Web: Discovering Knowledge from Hypertext Data Soumen Chakrabarti Morgan Kaufmann Publishers
3. DATA MINING THE WEB Uncovering Patterns in Web Content, Structure, and Usage, Zdravko Markov And Daniel T. Larose, Wiley-Interscience A John Wiley & Sons, Inc., Publication
4. Data Mining: Concepts and Techniques J. Han, M. Kamber Morgan Kaufmann

Reference Books:

1. Mining the Web: Discovering Knowledge from Hypertext Data, by Soumen Chakrabarti, Morgan Kaufmann Pub.

MSCDS213		2. Deep Learning		DP
DSE	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To understand the fundamentals and significance of Deep Learning and Artificial Neural Networks.
- To learn the architecture and training of neural networks using backpropagation techniques.
- To study CNNs, RNNs, LSTMs, GANs, and Autoencoders for advanced deep learning applications.
- To develop practical skills in designing, training, and evaluating deep learning models using TensorFlow and Keras.

Course Outcomes

After successful completion of the course, students will be able to:

- Explain the concepts, architectures, and learning methods of deep learning models.
- Design and implement neural networks for image, sequential, and predictive data analysis.
- Apply optimization, regularization, and feature enhancement techniques to improve model performance.
- Develop deep learning applications using frameworks such as TensorFlow and Keras for real-world problems.

UNIT-I:

Introduction To Deep Learning: Historical Trends in Deep Learning, Why DL is Growing, Artificial Neural Network, Non-linear classification example using Neural Networks: XOR/XNOR, Single/Multiple Layer Perceptron, Feed Forward Network, Deep Feed-forward networks, Stochastic Gradient –Based learning, Hidden Units, Architecture Design, Back- Propagation.

UNIT-II:

Convolution Neural Network (CNN): Introduction to CNNs and their applications in computer vision, CNN basic architecture, Activation functions-sigmoid, tanh, ReLU, Softmax layer, Types of pooling layers, Training of CNN in TensorFlow, various popular CNN architectures: VGG, Google Net, ResNet etc, Dropout, Normalization, Data augmentation.

UNIT-III

Recurrent Neural Network (RNN): Introduction to RNNs and their applications in sequential data analysis, Back Propagation Through Time (BPTT), Vanishing Gradient Problem, gradient clipping Long Short-Term Memory (LSTM) Networks, Gated Recurrent Units, Bidirectional LSTMs, Bidirectional RNNs.

UNIT- IV

Generative Adversarial Networks (GANS): Generative models, Concept and principles of GANs, Architecture of GANs (generator and discriminator networks), Comparison between discriminative and generative models, Generative Adversarial Networks (GANs), Applications of GANs. Auto-encoders, Architecture and components of auto-encoders (encoder and decoder), Training an auto-encoder for data compression and reconstruction. Transformer-Based Models: Introduction to Transformer Architecture, Attention Mechanism, Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer (GPT), Comparison of BERT and GPT, Applications in Natural Language Processing and Generative AI.

Text Books:

1. Deep Learning: An MIT Press Book by Ian Goodfellow and Yoshua Bengio Aaron courville.
2. Michael Nielson, Neural Networks and Deep Learning, Determination Press,2015.
3. Satish Kumar, Neural networks: A classroom Approach, Tata McGraw-Hill Education, 2004

Reference Books:

1. Deep Learning with Python, Francois Chollet, Manning publications, 2018.
2. Advanced Deep Learning with Keras, Rowel Atienza, PACKT Publications, 2018
3. Generative Deep Learning Teaching Machines to Paint, Write, Compose, and Play by David Foster Foreword by Karl Friston.
4. Natural Language Processing with Transformers Building Language Applications with Hugging Face by Lewis Tunstall, Leandro von Werra, and Thomas Wolf Foreword by Aurélien Géron O'Reilly

MSCDS214		Optimization Techniques Lab		OT-PR1
DSC	Credits: 2	HPW: 4	CA: 20	ESE: 30

List of Experiments:
Programs List using Conventional and TORA

1. Mathematical formulation of LPP.
2. Graphical method of solving an LPP.
3. Simplex method and Big-M method.
4. Two-phase simplex method.
5. Solution of LPP using dual.
6. Dual simplex method.
7. Optimum solution of balanced and unbalanced Transportation problems.
8. Solution of Assignment problems
9. Solution of sequencing problems of n-jobs with m ($m \geq 2$) machines.
10. Queueing problems
11. Solution of Games.
12. Net-Works.

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

MSCDS215		1. Web Mining Lab		WM-PR2
DSE	**Credits: 2**	**HPW: 4**	**CA: 20**	**ESE: 30**

List of Experiments:

1. Write a program to remove stop words from an input text document.
   Note the compression achieved
2. Implement a program to classify text
3. Write a program to simulate a web crawler
4. Write a program to form clusters of consumers based on their buying patterns
5. Write a program to simulate a recommender system for an online shopping site

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

MSCDS215		2. Deep Learning Lab		DP-PR2
DSE	Credits: 2	HPW: 4	CA: 20	ESE: 30

1. a. Design a single unit perceptron for classification of a linearly separable binary dataset without using pre-defined models. Use the Perceptron() from sklearn.
   b. Identify the problem with single-layer perceptron. Classify using AND, OR, and XOR datasets and analyze the result.
2. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets. Vary the activation functions used and compare the results.
3. Build a Deep Feed Forward ANN by implementing the Backpropagation algorithm and test the same using appropriate data sets. Use the number of hidden layers >=4.
4. Design and implement an Image classification model to classify a dataset of images using Deep Feed Forward NN. Record the accuracy corresponding to the number of epochs. Use the MNIST, CIFAR-10 datasets.
5. Design and implement a CNN model (with 2 layers of convolutions) to classify multi category image datasets. Record the accuracy corresponding to the number of epochs. Use the MNIST, CIFAR-10 datasets.
6. Design and implement a CNN model (with 4+ layers of convolutions) to classify multi category image datasets. Use the MNIST, Fashion MNIST, CIFAR-10 datasets. Set the No. of Epoch as 5, 10 and 20. Make the necessary changes whenever required. Record the accuracy corresponding to the number of epochs. Record the time required to run the program, using CPU as well as using GPU in Colab
7. Design and implement a CNN model (with 2+ layers of convolutions) to classify multi category image datasets. Use the concept of padding and Batch Normalization while designing the CNN model. Record the accuracy corresponding to the number of epochs. Use the Fashion MNIST/MNIST/CIFAR10 datasets.
8. Use the concept of Data Augmentation to increase the data size from a single image.
9. Implement RNN for sentiment analysis on movie reviews.
10. Implement Bidirectional LSTM for sentiment analysis on movie reviews.

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

TELANGANA UNIVERSITY,
NIZAMABAD
FACULTY OF SCIENCE
M.SC. (DATA SCIENCE) COURSE CURRICULUM

SEMESTER – IV (22 CREDITS)

S.No.	Type	Title of the Paper	Credits	HPW	CA	ESE	Marks
1	DSC	Natural Languages Processing	4	4	40	60	100
2	DSC	IoT And Predictive Analytics	4	4	40	60	100
3	DSE	1. Cloud Computing 2. Big Data	4	4	40	60	100
4	DSC-PR-1	Natural Languages Processing Lab	2	4	20	30	50
5	DSE-PR-2	1. Cloud Computing Lab 2. Big Data Lab	2	4	20	30	50
6	RP	Research Project (Report (100) + Viva (50)	**6**	6	-	150	150
	For Research Project, a group of six students will be guided for 2 HPW, so the total workload will be the multiple of number of groups and 2 HPW						
Total			**22**	**26**	**140**	**410**	**550**

DSC: Discipline Specific Course
DSE: Discipline Specific Elective
RP: Research Project
PR=Practical

MSCDS221		Natural Languages Processing		NLP
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To introduce the fundamental concepts, techniques, and applications of Natural Language Processing (NLP).
- To understand linguistic phenomena and model them using formal grammars and statistical methods.
- To develop skills in designing, training, and evaluating NLP systems using supervised and unsupervised learning techniques.
- To learn the design and implementation of NLP algorithms and language modeling techniques.

Course Outcomes

After successful completion of the course, students will be able to:

- Analyze linguistic structures and model them using formal grammar techniques.
- Apply statistical and probabilistic methods for NLP tasks and language modeling.
- Design, implement, and evaluate NLP algorithms using appropriate training methodologies.
- Develop NLP systems using supervised and unsupervised learning approaches for real-world applications.

UNIT –I:

Natural Language processing (NLP) : Introduction, Applications or Use cases of NLP, Components of NLP, Steps in NLP, Finding the Structure of Words: Words and Their Components, Lexemes, Morphemes, Morphology, Problems in morphological processing, Typology, Morphological Typology, Natural Language Processing with python NLTK package (Text Preprocessing Tasks): Word Tokenization, Sentence Tokenization, Filtering Stop words, Stemming, Tagging Parts of Speech, Lemmatization, Chunking, Chinking, Named Entity Recognition, Term Frequency and Inverse Document Frequency (TF-IDF).

UNIT–II:

Syntax Analysis: Parsing Natural Language, Tree banks: A Data-Driven Approach to Syntax, Representation of Syntactic Structure: Syntax Analysis using Dependency Graph, Syntax Analysis using Phrase Structure Trees, Parsing Algorithms: Shift Reduce Parsing, Hyper Graphs and Chart Parsing (CYK Parsing), Models for ambiguity Resolution in Parsing: Probabilistic Context Free Grammar, Generative Models, Discriminative models for Parsing.

UNIT-III:

Language Modeling: Introduction, N-Gram Models, Language Model Evaluation, Parameter Estimation, Language Model Adaptation, Types of Language Models, Language-Specific Modeling Problems.

UNIT–IV:

Semantic Parsing: Introduction, Semantic Interpretation, System Paradigms, Word Sense Systems, Software. Predicate-Argument Structure, Meaning Representation Systems, Software. Discourse Processing: Cohesion, Reference Resolution, Discourse Cohesion and Structure.

Text Books:

1. Multilingual natural Language Processing Applications: From Theory to Practice–Daniel M.Bikel and Imed Zitouni, Pearson Publication.
2. Speech and Natural Language Processing-Daniel Jurafsky & James H Martin, Pearson Publications.

Reference Books:

1. Natural Language Processing and Information Retrieval: Tanvier Siddiqui, U.S.Tiwary.

MSCDS222		IoT And Predictive Analytics		IPA
DSC	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To understand the fundamental concepts, architecture, and applications of the Internet of Things (IoT).
- To learn methods for collecting, storing, and processing IoT-generated data.
- To apply predictive analytics and machine learning techniques on IoT data.
- To develop IoT-based solutions while exploring IoT platforms, security, ethics, and future trends.

Course Outcomes

After successful completion of the course, students will be able to:

- Explain IoT architecture, components, and communication protocols.
- Collect, preprocess, and manage IoT data for analytics applications.
- Apply predictive models and machine learning techniques for forecasting and anomaly detection.
- Evaluate IoT platforms, security challenges, and emerging trends in IoT analytics.

Unit I

Introduction to IoT Definition and Evolution of IoT – IoT Ecosystem – IoT Architecture – Physical and Logical Design – IoT Enabling Technologies – Sensors and Actuators – Communication Protocols (Zigbee – Bluetooth – MQTT – HTTP – CoAP) – IoT Applications in Data Analytics.

Unit II

IoT Data Acquisition and Management Data Collection from IoT Devices – Streaming Data and Time-Series – Edge vs. Cloud Computing – Data Storage Solutions (InfluxDB – MongoDB – AWS IoT Core) – RealTime Data Processing – Sensor Data Characteristics – Preprocessing IoT Data for Analytics

Unit III

Introduction to Predictive Analytics: Definition and Importance – Types of Predictive Models – Supervised Learning Techniques – Linear and Logistic Regression – Decision Trees – Random Forests – Time Series Forecasting – Model Evaluation: MAE – RMSE – R^2 – Cross-Validation.

Unit IV

Predictive Analytics on IoT Data Data Integration from IoT Sources – Feature Engineering for Sensor Data – Predictive Maintenance – Anomaly Detection – Forecasting with IoT Data (ARIMA – LSTM – Prophet) – Visualization and Dashboarding – Case Studies in Smart Cities, Healthcare, and Industry 4.0

IoT Platforms and Future Trends: Overview of IoT Platforms (Thing Speak – AWS IoT – Google Cloud IoT – Azure IoT Hub) – Security and Privacy in IoT Data – Ethics in Predictive Analytics

Text Books:

1. Arshdeep Bahga & Vijay Madisetti, Internet of Things: A Hands-On Approach, Techno Series, 2023.
2. Eric Siegel, Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, Wiley, 2016

MSCDS223		1. Cloud Computing		CC
DSE	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To introduce the fundamental concepts and architecture of cloud computing and virtualization.
- To understand cloud service models, deployment models, and migration techniques.
- To develop knowledge of virtualization technologies and cloud resource management.
- To study cloud security, privacy, and service delivery mechanisms in cloud environments.

Course Outcomes

After successful completion of the course, students will be able to:

- Explain cloud computing service models, deployment methods, and virtualization concepts.
- Apply cloud programming and deployment techniques in real-world environments.
- Analyze the advantages, limitations, and security issues of virtualization and cloud computing.
- Evaluate authentication, confidentiality, privacy, and security implications in cloud systems.

Unit - I

Introduction to Computing: Computing Paradigms: Parallel Computing, Distributed Computing, Cluster Computing, Grid Computing, Cloud Computing Features of cloud and grid platforms, parallel and distributed programming paradigms Benefits and need for Cloud Computing, Cloud Services Requirements - Cloud and Dynamic Infrastructure NIST model - Cloud Computing Characteristics -. Cloud Models**:** Cloud deployment Models, Characteristics - cloud service models, Principal Technologies - Cloud Strategy Cloud Design and Implementation using SOA - Service Level Agreements SLA- Conceptual Cloud Model

Unit – II

Cloud Offers Management and Solutions: Cloud Ecosystem - Business and IT Perspective, Cloud Business Process Management - Cloud Service Management - Cloud Stack - Computing on Demand (CoD) – Cloud Offerings: Information Storage, Retrieval, Archive and Protection - Cloud Analytics, Testing under Cloud - Information Security - Storage Cloud. Cloud Management: Resiliency – Provisioning - Asset Management - Cloud Governance - High Availability and Disaster Recovery

Unit – III

Cloud Virtualization Technology: Virtualization Defined - Virtualization Benefits –Cloud Server Virtualization - Virtualization for x86 Architecture - Hypervisor Management Software - Logical Partitioning (LPAR)- VIO Server - Virtual Infrastructure Requirements. Cloud Virtualization: Storage virtualization - Storage Area Networks - Virtualized Data Center.

Unit-IV

Cloud Service Providers: Google, Cloud Platform, Cloud Storage, Google Cloud Connect, Google Cloud Print, Google App Engine, Amazon Web Services, Amazon Elastic Compute Cloud, Amazon Simple Storage Service, Amazon Simple Queue, service, Microsoft, Windows Azure, Microsoft Assessment and Planning Toolkit. SharePoint.

Text Book:

1. Essentials of cloud Computing: K. Chandrasekhran, CRC press, 2014

Reference Books:

1. Cloud Computing: Principles and Paradigms by Rajkumar Buyya, James Broberg and Andrzej M. Goscinski, Wiley, 2011.
2. Distributed and Cloud Computing, Kai Hwang, Geoffery C. Fox, Jack J. Dongarra, Elsevier, 2012.
3. Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, Tim Mather, Subra Kumaraswamy, Shahed Latif, O'Reilly, SPD, rp 2011.

MSCDS223		2. Big Data		BD
DSE	Credits: 4	HPW: 4	CA: 40	ESE: 60

Course Objectives

- To understand the concepts, platforms, and applications of Big Data.
- To learn Apache Hadoop architecture, HDFS, and Hadoop ecosystem components.
- To develop skills in processing distributed data using MapReduce and Hadoop tools.
- To apply analytics techniques on structured and unstructured Big Data.

Course Outcomes

After successful completion of the course, students will be able to:

- Identify Big Data concepts and their business applications.
- Explain Hadoop architecture and components of the Hadoop ecosystem.
- Access, process, and manage data in distributed file systems using Hadoop.
- Develop and analyze Big Data solutions using Hadoop and related technologies.

Unit I

Introduction: What is big data, why big data, convergence of key trends, unstructured data, industry examples of big data, web analytics, big data and marketing, fraud and big data, risk and big data, credit risk management, big data and algorithmic trading, big data and healthcare, big data in medicine, advertising and big data, big data technologies, introduction to Hadoop, open source technologies, cloud and big data, mobile business intelligence, Crowd sourcing analytics, inter and trans firewall analytics.

Unit II

Hadoop: Data format, analyzing data with Hadoop, scaling out, Hadoop streaming, Hadoop pipes, design of Hadoop distributed file system (HDFS), HDFS concepts, Java interface, data flow, Hadoop I/O, data integrity, compression, serialization, Avro, file-based data structures.

Unit III

MapReduce: MapReduce workflows, unit tests with MRUnit, test data and local tests, anatomy of MapReduce job run, classic Map-reduce, YARN, failures in classic Map-reduce and YARN, job scheduling, shuffle and sort, task execution, MapReduce types, input formats, output formats.

Unit IV

Big Data Analysis: Hbase, data model and implementations, Hbase clients, Hbase examples, praxis. Cassandra, Cassandra data model, Cassandra examples, Cassandra clients, Hadoop integration, Hive, data types and file formats, HiveQL data definition, HiveQL data manipulation, HiveQLqueries.

Text Books

1. Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning, Raj kamal,PreetiSaxena, McGraw Hill, 2018.
2. Big Data, Big Analytics: Emerging Business intelligence and Analytic trends for Today's Business, Michael Minelli, Michelle Chambers, and AmbigaDhiraj, John Wiley & Sons,2013.

Reference Books

1. Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013
2. Hadoop: The Definitive Guide, Tom White, Third Edition, O'Reilley, 2012.
3. Hadoop Operations, Eric Sammer, O'Reilley, 2012.
4. Programming Hive, E. Capriolo, D. Wampler, and J. Rutherglen, O'Reilley, 2012.
5. HBase: The Definitive Guide, Lars George, O'Reilley, 2011.
6. Cassandra: The Definitive Guide, Eben Hewitt, O'Reilley, 2010. 7. Programming Pig, Alan Gates, O'Reilley, 2011.

MSCDS224		Natural Languages Processing Lab		NLP-PR1
DSC	Credits: 2	HPW: 4	CA: 20	ESE: 30

List of Experiments:

01.
   a) Write a python program to perform tokenization by word and sentence using nltk.
   b) Write a python program to eliminate stopwords using nltk.
   c) Write a python program to perform stemming using nltk.

02.
   a) Write a python program to perform Parts of Speech tagging using nltk.
   b) Write a python program to perform lemmatization using nltk.

03.
   a) Write a python program for chunking using nltk.
   b) Write a python program to perform Named Entity Recognition using nltk

04.
   a) Write a python program to find Term Frequency and Inverse Document Frequency (TF-IDF).
   b) Write a python program for CYK parsing (CockeYounger-Kasami Parsing) or Chart Parsing.

05.
   a) Write a python program to find all unigrams, bigrams and trigrams present in the given corpus.
   b) Write a python program to find the probability of the given statement "This is my cat" by taking the an example corpus into consideration

06. Use the Stanford named Entity recognizer to extract entities from the documents. Use it programmatically and output for each document which named entities it contains and of which type

07. Choose any corpus available on the internet freely. For the corpus, for each document, count how many times each stop word occurs and find out which are the most frequently occurring stop words. Further, calculate the term frequency and inverse document frequency as The motivation behind this is basically to find out how important a document is to a given query. For e.g.: If the query is say: "The brown crow". "The" is less important. "Brown" and "crow" are relatively more important. Since "the" is a more common word, its tf will be high. Hence we multiply it by idf, by knowing how common it is to reduce its weight.

08. Write the python code to perform sentiment analysis using NLP.

09. Write the python code to develop Spam Filter using NLP.

10. Write the python code to detect Fake News using NLP.

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

CDS225		1. Cloud Computing Lab		CC-PR2
DSE	Credits: 2	HPW: 4	CA: 20	ESE: 30

List Of Experiments:

1. Installation and configuration of Virtual Machine using VMware.
2. Study and Implementation of Infrastructure as a Service.
3. Installation and Configuration using Microsoft Azure Virtual Machine.
4. Install Google App Engine. Create hello world app and other simple web applications using Python/Java.
5. Create an Amazon EC2 Instance and Set up a Web Server on the Instance and Associate IP Address with the Instance.
6. Create a Database Instance in the Cloud using RDS.
7. Create a Database Instance in the Cloud using Google Cloud SQL.
8. Register with AWS and Create Windows/Linux Instance.
9. Create a S3 Storage Bucket and Store documents in Bucket.
10. Create a Static Web Hosting on S3 with name Kakatiya university.
11. Build a Serverless Web Application on AWS Cloud to GET KU employee details with emailid only.
12. Build a Serverless Web Application on AWS Cloud to POST KU employee details with emailid only.

Note:

- All the concepts of syllabus and exercises from Text Book must be translated into programs which must be practiced, executed and written down in the practical record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory

MSCDS225		2. Big Data Lab		BD-PR2
DSE	Credits: 2	HPW: 2	CA: 20	ESE: 30

Note:

- All the concepts of syllabus and exercises from Text Book must be translated into programs which must be practiced, executed and written down in the practical record book.

1. Install Hadoop and Implement the following file management tasks in Hadoop: Adding files and directories Retrieving files Deleting files and directories. 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
2. Develop a MapReduce program to implement Matrix Multiplication
3. Develop a Map Reduce program that mines weather data and displays appropriate messages indicating the weather conditions of the day.
4. Develop a MapReduce program to find the tags associated with each movie by analyzing movie lens data.
5. Implement Functions: Count – Sort – Limit – Skip – Aggregate using MongoDB
6. Write Pig Latin scripts to sort, group, join, project, and filter the data.
7. Use Hive to create, alter, and drop databases, tables, views, functions, and indexes.
8. Implement a word count program in Hadoop and Spark.
9. Use CDH (Cloudera Distribution for Hadoop) and HUE (Hadoop User Interface) to analyze data and generate reports for sample datasets

NOTE:

- all the concepts of the syllabus and exercises from the textbook must be translated into programs that must be practiced, executed, and written down in the record book.
- In the external lab examination, the student has to compile and execute at least two programs.
- External Viva-voce is compulsory.

MSCDS226		Research Project (Report (100) + Viva (50)		RP
RP	Credits: 6	HPW: 6	CA: --	ESE: 150

Note:

- The project work should be carried out by the Student groups and the number of students in each group should not exceed three.
- Two Seminars must be delivered by all the student groups before the final project viva, one on the design part and the other on the implementation.
- The Project work, being the vital component of this professional program, needs to be carried out with due care and dedication by all the student groups.
- The project work is not just the partial fulfilment of the course requirements, but it provides a mechanism to demonstrate the ASK (Attitude, Skills, and knowledge) elements with specialization.
- The students are expected to work on a real-time project on latest platforms preferably in some industrial UNIT/ R&D Laboratories / Educational Institution / Software Company.
- Students are encouraged to work in their interested area.
- The student groups can formulate a project problem by a thorough interaction with his / her Guide of the concerned college.
- Approval of the project proposal is mandatory by his/her Guide, then only the student group is expected to commence working on and complete it.
- The student groups must make use of the latest software packages for the development of the project.
- The problem domain and the specifications chosen should be genuine and feasible for the implementation.