M.Sc. (DATA SCIENCE) COURSE CURRICULUM

(W.E.F. THE ACADEMIC YEAR 2025-26)

Approved by

BOARD OF STUDIES

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

TELANGANA UNIVERSITY,

NIZAMABAD

M.SC. (DATA SCIENCE) COURSE CURRICULUM

(W.E.F. ACADEMIC YEAR 2025-2026)

I-SEMESTER

Paper Code	Title of the paper	Credits	No of Hours per week	Total Marks
Theory Papers				
MDS-101	Mathematical Foundations for Data Science.	3	3	100
MDS-102	Design and Analysis of Algorithms.	3	3	100
MDS-103	Java Programming.	3	3	100
MDS-104	Statistical Inference.	3	3	100
Practical Papers				
MDS-105	Statistical Process for Data Science using R-programming.	2	4	50
MDS-106	Design and Analysis of Algorithms Lab using Python.	2	4	50
MDS-107	Java Programming Lab.	2	4	50
MDS-108	Statistical Inference using Python.	2	4	50
	Total	**20**	**12+16**	**600**

II-SEMESTER

Paper Code	Title of the paper	Credits	No of Hours per week	Total Marks
Theory Papers				
MDS-201	Advanced Machine Learning Techniques	3	3	100
MDS-202	Artificial Intelligence	3	3	100
MDS-203	Optimization Techniques	3	3	100
MDS-204	Software Engineering	3	3	100
Practical Papers				
MDS-206	Advanced Machine Learning Techniques using Python Lab	2	4	50
MDS-207	Artificial Intelligence Lab	2	4	50
MDS-208	Optimization techniques using R	2	4	50
MDS-209	Data Handling using SPSS	2	4	50
	Total	**20**	**24**	**600**

III-SEMESTER

Paper Code	Title of the paper	Credits	No of Hours per week	Total Marks
Theory Papers				
MDS-301	Deep Learning Techniques	3	3	100
MDS-302	Computer Networks	3	3	100
MDS-303 (E-I)	a) Text Data Analytics b) Statistical Pattern Recognition c) Cloud Computing	3	3	100
MDS-304 (E-II)	a) Big Data Analytics b) Image Processing & Analysis c) Web technologies	3	3	100
Practical Papers				
MDS-305	Deep Learning Techniques & Computer Networks using Python	2	4	50
MDS-306	E-I & E-II using Python	2	4	50
MDS-307	Mini Project	4	8	100
	No of Credits	**20**	**12+16**	**600**

IV-SEMESTER

Paper Code	Title of the paper	Credits	No of Hours per week	Total Marks
Theory Papers				
MDS-401	Cryptography	3	3	50
MDS-402	Data Mining	3	3	50
MDS-403 E-III	a) Sentimental Analysis b) Computer Vision c) Machine Learning Operations	3	3	50
MDS-404 E-IV	a) Data Stream mining b) Web Mining c) Scalable Architecture	3	3	50
Practical Papers				
MDS-405	Cryptography & DM	2	4	50
MDS-406	E-I & E-II	2	4	50
MDS-407	Capstone Project	4	8	100
	No of Credits	20	12+16	**600**

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-101-T: PAPER- I: MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE

UNIT – I

Linear Algebra: Vector spaces, Subspaces, Basis and dimension of a vector space, linear dependence and independence, spanning set. Linear transformation, kernel, range, Matrix Representation of a linear transformation, Matrices: Trace and Rank of a Matrix and their properties, Determinants, Inverse, symmetric, orthogonal and idempotent matrices and their properties, Gauss elimination, row canonical form, diagonal form, triangular form and its Applications, Characteristic roots and vectors, Statement of Caley-Hamilton theorem and its applications, Orthogonal and Spectral decomposition of a real symmetric matrix, Singular value Decomposition.

UNIT – II

Combinatorics: Basic counting principle, inclusion-exclusion for two-sets, pigeonhole principle, permutations and combinations, Binomial coefficient and identities, generalized permutations and combinations. principle of inclusion-exclusion, applications of inclusion-exclusion. Recurrence Relations: introduction, solving linear recurrence relations, generating functions. Real time applications of combinatorial concepts.

UNIT-III

Graph Theory: Basic concepts to Graphs; Isomorphic graphs and simple problems; Trees: definitions, properties, and simple problems, tree traversals, spanning tree constructions: breadth and depth first search, minimal spanning tree constructions: Kruskal's algorithm. Planar and Hamiltonian graphs and simple problems and applications; Graph coloring and its applications. Real time applications of graphical concepts.

REFERENCE BOOKS

1. Gilbert Strang (2016): Introduction to linear algebra, 5/e., Wellesley-Cambridge.
2. David C. Lay (2019): Linear Algebra and Its Applications, Pearson, 5/e.
3. Joe L. Mott, Abraham Kandel, Theoder P. Baker, Discrete Mathematics for Computer Scientists and Mathematicians

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-102-T: PAPER- II: DESIGN AND ANALYSIS OF ALGORITHMS

UNIT I

Introduction to Algorithms: Algorithm Specification, Performance Analysis, Randomized Algorithms. **Elementary Data Structures**: Stacks and Queues, Trees, Dictionaries, Priority Queues, Sets and Disjoint Set Union, Graphs.

Divide and Conquer: Binary Search, Finding the Maximum and Minimum, Merge Sort; Quick Sort, Selection sort, Strassen's Matrix Multiplication, Convex Hull.

Greedy Method: Knapsack Problem, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees (Kruskal's & Prim's), Single Source Shortest Paths (Dijkstra's).

UNIT-II

Dynamic Programming: General Method, Multistage Graphs, All-Pairs Shortest Paths, Single-Source Shortest Paths, Optimal Binary Search Trees, 0/1 Knapsack, Traveling Salesperson Problem.

Back Tracking technique: General Method, 8-Queens Problem, Sum of Subsets, Graph Colouring, Hamiltonian Cycles, Knapsack Problem.

UNIT -III

Branch-Bound technique: General Method, 0/1 Knapsack Problem, Traveling Sales Person problem.

NP-Hard and NP-Complete Problems: Basic Concepts, Cook's Theorem, NP-Hard and NP-Complete problems. Graph Problems, NP-Hard Scheduling Problems, NP-Hard Code Generation, Some Simplified NP-Hard Problems.

REFERENCE BOOKS

1. E Horowitz, S Sahni, S Rajasekaran, (2007): Fundamentals of Computer Algorithms, 2/e, Universities Press.
2. T.H. Cormen, C.E. Leiserson, R.L Rivert, C Stein, (2010): Introduction to Algorithms, 3/e, PHI.
3. R. Pannerselvam (2007): Design and Analysis of Algorithms, PHI.
4. Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-103-T: PAPER- III: JAVA PROGRAMMING

UNIT – I

Java Programming Fundamentals: Introduction, Overview of Java, Data Types, Variables and Arrays, Operators, Control statements, Classes, Methods, Inheritance, Packages and Interfaces. I/O basics: Byte & Character Streams, Reading Console input and output, Scanner Class, Console Class, Print Writer Class, String Handling, Exception Handling, Multithreaded Programming. Overview of Networking: Working with URL, connecting to a Server, Implementing Servers, serving multiple Clients, Sending E-Mail, Socket Programming, Internet Addresses, URL Connections. AWT: Introduction, AWT Class Hierarchy, Creating Container, Adding Components, Layout, Using Panel, TextField, TextArea, List, Checkbox, CheckBoxGroup, Choice, EventHandling, DialogBoxes, ScrollBar, Menu.

UNIT – II

Swing: Containment Hierarchy, Adding Components, JTextField, JPasswordField, JTable, JComboBox, JProgressBar, JList, JTree, JColorChooser, Dialogs. Java Database Connectivity (JDBC): Introduction, JDBC Drivers, JDBC Architecture, JDBC Classes and Interfaces, Loading a Driver, Making a Connection, Execute SQL Statement, SQL Statements, Retrieving Result, Getting Database Information, Scrollable and Updatable Resultset, ResultSet Metadata.

UNIT – III

Servlet: Introduction to Servlet, Servlet Life Cycle, advantages, Sharing Information, Initializing a Servlet, Writing Service Methods, Filtering Requests and Responses, Invoking Other Web Resources, Accessing the Web Context, Maintaining Client State, Finalizing a Servlet. Java Server Pages: Introduction to JSP, JSP Engine, Anatomy of a JSP Page, JSP Syntax, Life Cycle of a JSP Page, Creating Static Content, Creating Dynamic Content. .

REFERENCE BOOKS

1. Uttam K. Roy, Advanced Java programming
2. Herbertt Schildt, Java Complete Reference
3. Cay S. Horstmans, Gray Coronell, Core Java Vol. II – Advanced Features
4. Sharanam Shah, Vaishali Shah, Java EE 7 for Beginners

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-104-T: PAPER- IV: STATISTICAL INFERENCE

UNIT-I

Estimation Theory: Basic concepts to estimation; Concepts, examples, applications and simple problems on criteria for good estimator: Unbiasedness, consistency, efficiency and sufficiency, Cramer-Rao inequality, Rao-Blackwell theorem, Fisher Information, Lehmann-Scheffe theorem. Simple Problems on UMVUE.

Methods of Estimation: Method of Moments, Least squares and Maximum Likelihood, Properties and Simple problems. Resampling methods: Jackknife, Bootstrap, 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-II

Testing of Hypotheses: Neyman-Pearson Lemma, Most Powerful tests, Uniformly Most Powerful tests, Likelihood ratio tests, Sequential Probability Ratio Tests.

Non parametric tests: One and two sample tests (Kolmogorov Smirnov, Kruskal Wallis & Friedman test, Kendal's tau, Ansari broadly tests)

UNIT-III

Non-parametric Density Estimation: Rosenblatt's naïve density estimator, its bias and variance. Consistency of Kernel density estimators and its MSE.

Simulation: Introduction, generation of random numbers for Uniform, Normal, Exponential, Cauchy and Poisson Distributions. Estimating the reliability of the random numbers. Priori and Posteriori distributions, conjugate families, Bayesian estimation of parameters, MCMC algorithms: Metropolis Hasting and Gibbs Sampler.

REFERENCES

1. Rohatgi, V.K.: An Introduction to Probability Theory and Mathematical Statistics (Wiley)
2. Gibbons: Non-Parametric Statistical Inference, (TMH)
3. Lehman, 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)

Instruction to Practical's of M.Sc. Data Science

1. The semesters I & II, each has **Four practical papers,** each has **two credits** with **4 hours of lab** with weightage of **50 marks each**.
2. **Each student has to spend minimum 60 hours** in lab for each practical paper and has to practice on various data sets available in various web sources.
3. Each practical record should contain all practical's mentioned in the syllabus and maintaining of **Practical records is mandatory** (submission at the time examination) irrespective of existence or non-existence of practical record marks allocation.
4. The **statistical analysis report should follow the steps of Data Analysis** depends on the practical (Problem, Data set considered (Data source), Data description, Data objectives, hypothesis framed on population, statistical techniques applicable (as per syllabus of paper), writing of base program code to familiar with computational procedure rather than usage of packages), outputs and results, data interpretation, conclusions on the data set).
5. Each practical record should be written with own hand writing (not computer printouts) and should be taken the signature of the concerned faculty with date of the practical done.
6. The Semester end practical exam question paper contains **answer any two out of the three** questions given in **2 hours duration** (including its implementation)
7. **The practical examination question paper is also common to all the students** of all colleges and conducted and scheduled by Head, Department of Statistics, O.U., Hyderabad, with the appointment of external examiners.
8. **All answers to the questions should be written in the practical answer booklets at the time of practical examination** (including executed program, output, interpretation and conclusions and incase of data analysis all the steps should be written as mentioned above.

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-105-P: PAPER- V (PRACTICAL-I)

STATISTICAL PROCESS FOR DATA SCIENCE USING R

List of Practical's:

1. Basics of R-programming and R-studio for data handling.
2. Data understanding, data description, Measurement of scales, Data objectives, Formation of Hypothesis's. Sequential steps for writing data analysis report.
3. Evaluation of Data pre-processing steps. Data transformations (Standardize, Normalize, converting data from one scale to other scales etc.).
4. Evaluation of Descriptive Statistics to various measurement of scales data sets.
5. Data Visualization using R: Drawing One dimensional diagram (Pictogram, Pie Chart, Bar Chart), two-dimensional diagrams (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot), other diagrammatical / graphical representations like, Gantt Chart, Heat Map, Box-Whisker Plot, Area Chart, Correlation Matrices.
6. Correlation Analysis (parametric and nonparametric), Simple and Multiple linear Regression model fitting and its analysis.
7. Testing of Hypothesis-I: Parametric tests (z-, χ^2, t-, F-tests, ANOVA).
8. Testing of Hypothesis-I: Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test).
9. Statistical analysis for qualitative data.
10. Applying the modelling process, Model evolution, over fitting, under fitting, cross validation concepts, Model Performance (train/test, K fold and leave out one approaches) for qualitative and Quantitative data.

Note: The implementation of the above list of practical's to be applied on the sample data sets available in various web sources and should be practiced by each student. For example, www.kaggle.com . contains thousands of data sets with different measurement of scales; few are: Fishers Iris Dataset; Online food dataset, Wine quality data set, water portability dataset, Heart data set, Protease data set, Mortagaze data set, flights dataset; Sustainable Development Data; Credit Card Fraud Detection; Employee dataset; Heart Attack Analysis & Prediction Dataset; Dataset for Facial recognition; Covid_w/wo_Pneumonia Chest Xray Dataset; Groceries dataset; Financial Fraud and Non-Fraud News Classification; IBM Transactions for Anti Money Laundering.

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-106-P: PAPER- VI (PRACTICAL-II):

DESIGN & ANALYSIS OF ALGORITHMS LAB USING PYTHON

List of Practical's using python programming (not usage of packages)

1. Divide and Conquer method of implementation of sorting/ searching data / data set using:
   (i) Selection Sort
   (ii) Merge Sort
   (iii) Quick Sort.
   (iv) Construction of Heap, Maintain heap and Heap sort.
   (v) Binary Search.
   (vi) Strassen's Matrix multiplication
2. Greedy method implementation
   (i) Fractional Knapsack problem.
   (ii) Job Sequencing with Deadlines,
   (iii) Minimum-Cost Spanning Trees (Kruskal's & Prim's),
   (iv) Single Source Shortest Paths (Dijkstra's).
3. Dynamic programming technique implementation for
   (i) Travelling sales person problem.
   (ii) Multistage Graph problem,
   (iii) All-Pairs Shortest Paths (Warshal),
   (iv) Single-Source Shortest Paths (Bellman ford),
   (v) Optimal Binary Search Trees.
4. Back tracking technique implementation for
   (i) 0-1 Knapsack problem.
   (ii) 8-Queens Problem,
   (iii) Hamiltonian Graph problem
5. Branch and bound Implementation for
   (i) 0-1 Knapsack problem.
   (ii) Travelling salesman problem.

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-107-P: PAPER- VII (PRACTICAL-III): JAVA PROGRAMMING LAB

List of Practical's using Java programming

1. Java programs to the Mathematical / Statistical applications to demonstrate core Java and OOPs concepts. (working with functions, classes, abstract class, interfaces, string handling and string buffer classes, user defined packages, event handling, exception handling, inheritance, polymorphism, multithreading, etc, Computation central tendencies, dispersions, moments, skewness, kurtosis, distributions, correlation and regression, Matrices)
2. Developing servlet applications on data bases. (like to accept H.T.No of a student from client and display the memorandum of marks from the server, Question-Answer Application using HttpServletRequest and HttpServletRequest interfaces etc).
3. Create JSP pages that prints (a) temperature conversion (from Celsius to Fahrenheit) chart; (b) current date and time (c). number of times page is referred after the page is loaded. simple JSP application to demonstrate the use of implicit object (at least 5). JSP Application to accept Registration Details from the user and store database table; Accept Registration Details from the user and store database table; Authenticate User Login as per the Registration Details. If Login Success then forward User to Index Page otherwise show Login failure Message; web Application to add items in the inventory using JSP.
4. Create GUI to present a set of choices for a user to select stationary products and display the price of Product after selection from the list; typical Editable Table which describing Employee for a software company; swing components using student registration form.
5. Create a Remote Object for simple arithmetic operators. Use AWT / SWING to create user interface.
6. Develop a Hibernate application to Store Feedback of Website Visitors in MySQL Database.
7. Write EJB applications using stateless session beans and state-full session beans.
8. Develop a Room Reservation System Application using Entity Beans.
9. Create Three-tire application using Servlets, JSP, EJB.

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, I-SEMESTER

MDS-108-P: PAPER- VIII (PRACTICAL-IV): STATISTICAL INFERENCE USING PYTHON

List of Practical's using Python programming

1. Data visualization: Diagrammatical / Graphical representation of the data in the form of dataset with different measurement of scales (Pictorial representation, Bar (simple, multiple, component, percent) and Pie Charts, Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot, Gantt Chart, Heat Map, Box - Whisker Plot, Waterfall Chart, Area Chart, Density Plot, Bullet Graph, Choropleth Map, Tree map, Path diagram, Network Diagram, Correlation Matrices).
2. Correlation and Regression Analysis (including simple (Pearsons and Spearman's), partial and multiple correlations, Simple, Multiple linear regression and logistic regression).
3. Parametric tests (z-, χ^2, t-, F-tests, ANOVA).
4. Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test, U-test, K-S test, Kruskal Wallis and Friedman test, Independence, goodness of fit, Kendal's tau, Ansari broadly tests).
5. Generation of Jackknife and Bootstrap samples and estimation of parameters and computation of bias.
6. Confidence Interval estimation for Binomial, Poisson, Normal and Exponential parameters.
7. Simulation: Generation of random numbers from various probability distributions (Uniform, Binomial, Poisson, Normal, Exponential, Gamma, Cauchy, Lognormal, and Weibull Distributions).
8. Bayesian estimation of parameters (using Metropolis Hasting and Gibbs Sampler).

M.SC. (DATA SCIENCE) COURSE CURRICULUM

(W.E.F. ACADEMIC YEAR 2024-2025)

M.SC. (DATA SCIENCE) I-YEAR, II-SEMESTER

Paper Code	Title of the paper	Credits	No of Hours per week	Total Marks
Theory Papers				
MDS-201	Advanced Machine Learning Techniques	3	3	100
MDS-202	Artificial Intelligence	3	3	100
MDS-203	Optimization Techniques	3	3	100
MDS-204	Software Engineering	3	3	100
Practical Papers				
MDS-206	Advanced Machine Learning Techniques using Python Lab	2	4	50
MDS-207	Artificial Intelligence Lab	2	4	50
MDS-208	Text Analytics Lab using Python	2	4	50
MDS-209	Data Handling using SPSS	2	4	50
	Total	**20**	**24**	**600**

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, II-SEMESTER

MDS-201: PAPER- I: ADVANCED MACHINE LEARNING TECHNIQUES

UNIT – I

Classification Techniques: Definitions, derivations, methods, properties, and applications of classifier algorithms: Logistic regression, Linear Discriminant Function (for binary outputs) with minimum mean squared error and using likelihood ratios (MVN populations), Bayes Mis-classification; Naïve Bayes classifier, Support Vector Machines, Decision Tree algorithms, Random Forest algorithm, Bootstrap algorithms (Bagging, Gradient, Ada- and XG-Boosting), KNN algorithm, Market-Basket Analysis.

UNIT – II

Multivariate Analysis Techniques: Definitions, goals, derivations, methods, properties, and applications of: Principal component analysis, Factor analysis, Multidimensional Scaling, Canonical Correlations and Canonical Variables, Conjoint Analysis, Path analysis, Correspondence analysis; Feature extraction and Feature selection techniques, Inter and intra class distance measures, Probabilistic distance measures.

UNIT - III

Cluster Analysis: Introduction, similarities and dissimilarities, Hierarchical and non-hierarchical clustering methods, Single, complete and average linkages, K-means, DBSCAN, CART, C4.5 methods.

Association Analysis: Problem definition, Frequent item set generation, Rule generation, Compact representation of frequent item sets, methods for generating frequent item sets, Aprori and FP-tree growth Algorithm.

REFERENCE BOOKS

1. Johnson, R.A, and Dean W. Wichern: Applied Multivariate Statistical Analysis.
2. Morrison, D: An Introduction to Multivariate Analysis.
3. Seber: Multivariate Observations
4. Anderson: An Introduction to Multivariate Analysis.
5. Bishop: Analysis of Categorical data.

SYLLABUS
M.SC. (DATASCIENCE) I-YEAR, II-SEMESTER

MDS-202-T: PAPER- II: ARTIFICIAL INTELLIGENCE

UNIT - 1

Introduction: History Intelligent Systems, Foundations of Artificial Intelligence, Sub areas of Al, Applications. Problem Solving - State - Space Search and Control Strategies: Introduction, General Problem-Solving Characteristics of problem, Exhaustive Searches, Heuristic Search Techniques, Iterative - Deepening A*, Constraint Satisfaction. Game Playing, Bounded Look - ahead Strategy and use of Evaluation Functions, Alpha Beta Pruning.

UNIT – II

Logic Concepts and Logic Programming: Introduction, Propositional Calculus Propositional Logic, Natural Deduction System, Axiomatic System, Semantic Table, A System in Propositional Logic, Resolution, Refutation in Propositional Logic, Predicate Logic, Logic Programming. Knowledge Representation: Introduction, Approaches to knowledge Representation, Knowledge Representation using Semantic Network, Extended Semantic Networks for KR, Knowledge Representation using Frames.

UNIT - III

Expert System and Applications: Introduction, Phases in Building Expert Systems Expert System Architecture, Expert Systems Vs Traditional Systems, Truth Maintenance Systems, Application of Expert Systems, List of Shells and tools. Uncertainty Measure - Probability Theory: Introduction, Probability Theory, Bayesian Belief Networks, Certainty Factor Theory, Dempster - Shafer Theory.

REFERENCE BOOKS

1. Saroj Kaushik, Artificial Intelligence, Cengage Learning India, First Edition, 2011.
2. Russell, Norvig, Artificial Intelligence: A Modern Approach, Pearson Education, 2nd Edition, 2004.
3. Rich, Knight, Nair, Artificial Intelligence, Tata McGraw Hill, 3rd Edition 2009.

SYLLABUS
M.SC. (DATASCIENCE) I-YEAR, II-SEMESTER
MDS-203: PAPER-III: OPTIMIZATION TECHNIQUES

UNIT–I

Meaning and scope of optimization techniques. Convex sets and their properties. General linear Programming Problem (LPP). Formulation of LPP. Statements of Fundamental theorem of LPP and other related theorems. Optimal Solution of LPP by graphical, Simplex and Charner's & two-phase methods. Concept of degeneracy and resolving it Concept of duality of LPP. Dual Primal relationship, Complementary slackness theorems, Fundamental Theorem of Duality. Dual simplex Algorithm.

UNIT-II

Transportation problem: Concept of transportation problem, TPP as a special case of LPP, Initial basic feasible solutions by North-West Corner Rule, Matrix minimum method and 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 and traveling salesman problem and its solution.

Problem of Sequencing: Optimal sequence of N jobs on two and three machines without passing.

UNIT-III

Integer Programming Problem: Gomory's cutting plane Algorithm for pure and mixed IPP Branch and bound Technique.

Networks: Basic concepts of Networks constraints; Construction of Network and critical path; PERT and CPM; Network flow problems. Time Cost Analysis.

Queuing Theory: Introduction, essential features of Queuing system, Operating characteristics of Queuing system (transient and steady states). Queue length, General relationships among characteristics. Probability distribution in queuing systems, distribution of Arrival and interarrival. Distribution of death (departure) process, service time. Classification of Queuing models and solution of Queuing models; M/M/1:∞/FIFO and M/M/1:N/FIFO

REFERENCES

1. Kantiswarup; Gupta P.K. and Singh, M.N.(1985) : Operations Research; Sultan Chand
2. Taha, H.A.(1982): Operations Research: An Introduction; MacMillan
3. Sharma,S.D.: Operations Research.
4. Hillier F.S. and Leiberman,G.J.(1962): Introduction to Operations Research; Holdon Day

SYLLABUS

M.SC. (DATASCIENCE) I-YEAR, II-SEMESTER

MDS-204-T: PAPER- IV: SOFTWARE ENGINEERING

UNIT – I

Software Engineering: The Nature of Software, Changing Nature of Software, Defining the Discipline, Software Process, Software Engineering Practice. The Software Process: A Generic Process Model, Defining a Framework Activity, Process Assessment and Improvement, Prescriptive Process Models, Specialized Process Models, Unified Process, Personal and Team Process Models. Defining Agility, Agile Process, Extreme Programming, Psychology of Software Engineering, Software Team Structures, Software Engineering Using the Cloud, Global Teams.

UNIT – II

Requirements: Core Principles of Modeling, Requirements Engineering, Establishing the Groundwork, Eliciting Requirements, Developing Use Cases, Building the Analysis Model, Requirements Analysis, UML Models That Supplement the Use Case, Identifying Analysis Classes, Specifying Attributes, Defining Operations, Class Responsibility-Collaborator Modeling, Associations and Dependencies, Analysis Packages. Design Concepts: Design within the Context of SE, Design Process, Design Concepts, Design Model, Software Architecture, Architectural Styles, Architectural Considerations, Architectural Design, Component, Designing Class-Based Components, Conducting Component-Level Design, Component-Based Development, User Interface Design Rules.

UNIT – III

Quality Management: Quality, Software Quality, Software Quality Dilemma, Achieving Software Quality, Defect Amplification and Removal, Reviews, Informal Reviews, Formal Technical Reviews, Elements of Software Quality Assurance, SQA Tasks, Goals, and Metrics, Software Reliability, A Strategic Approach to Software Testing, Test Validation Testing, System Testing, Debugging, Software Testing Fundamentals, White-Box Testing, Black-Box Testing, Path Testing, Control Structure Testing, Object-Oriented Testing Strategies & Methods, Security Engineering Analysis, Security Assurance, Security Risk Analysis.

REFERENCE BOOKS

1. Roger S Pressman, B R Maxim, Software Engineering – A Practitioner's Approach (8e)
2. Ian Sommerville: "Software Engineering".
3. Hans Van Vliet, Software Engineering.
4. D. Bell, Software Engineering for Students.
5. K.K. Aggarwal, Y. Singh, Software Engineering.
6. R. Mall, Fundamentals of Software Engineering

SYLLABUS

M.SC. (DATASCIENCE) I-YEAR, II-SEMESTER

MDS-205-P: PAPER- V (PRACTICAL-1):

ADVANCED MACHINE LEARNING USING PYTHON LAB

List of Practical's using Python

Note: preferably writing python code based on the procedure rather than usage of packages.

1. Implementation of classification techniques for the data sets and evaluation of its analysis using
    - (i) Logistic regression
    - (ii) Naïve Bayes classifier
    - (iii) Support Vector Machines,
    - (iv) Decision tree (ID-3) algorithms
    - (v) Random forest algorithm
    - (vi) Bagging and Boosting algorithms
    - (vii) KNN
2. Implementation of
    - (i) Principal component analysis
    - (ii) Factor analysis,
    - (iii) Multidimensional Scaling,
    - (iv) Path analysis.
3. Implementation of clustering using
    - (i) K-means,
    - (ii) DBSCAN,
    - (iii) CART,
    - (iv) C4.5 methods
4. Implementation of Association algorithms
    - (i) Aprori algorithm
    - (ii) FP-tree growth algorithm.

SYLLABUS

M.SC. (DATA SCIENCE) I-YEAR, II-SEMESTER

MDS-206-P: PAPER-VI (PRACTICAL-2):

ARTIFICIAL INTELLIGENCE USING R / PYTHON LAB

Programs List using R and Python

1. Implementation of A* and AO* algorithms.
2. Implementation of Alpha-beta pruning.
3. Implementation of search algorithms (BFS & DFS).
4. Implementation of Hill Climbing algorithm
5. Implementation of Gaming problems:
   (i) Tower of Hanoi problem
   (ii) Tic-Tac-Toe problem
   (iii) Water-Jug problem.
   (iv) 4-Queens problem.
   (v) 8 Puzzle problems.
   (vi) Monkey banana problem

SYLLABUS

M.SC. (DATASCIENCE) I-YEAR, II-SEMESTER

MDS-207-P: PAPER- VII (PRACTICAL-3):

OPTIMIZATION TECHNIQUES USING TORA & R

Programs List using R and TORA

1. Optimal Solution to a L.P.P. by
   (i) Graphical method
   (ii) Simplex method
   (iii) Charners Method
   (iv) Two-phase simplex method
   (v) Dual Simplex Method
   (vi) Duality
2. Transportation problem by MODI & stepping stone methods. (Balanced and unbalanced)
3. Assignment problem by Hungarian method (Balanced and unbalanced)
4. Traveling salesman problem by Hungarian method
5. Job sequencing problem for N jobs on 2 and 3 machines
6. Integer Programming Problem (Gomery's cutting plane and Branch & Bound methods)
7. Construction of Network diagram and finding Critical path using CPM and PERT.
8. Evaluation of Time cost analysis through CPM and PERT

SYLLABUS

M.SC. (DATASCIENCE) I-YEAR, II-SEMESTER

MDS-208-P: PAPER- VIII (PRACTICAL-4):

DATA HANDLING USING SPSS

List of Practical's

1. Basic operations of Data entry, Data import and export, I/O files handling etc.
2. **Data Visualization**: Pie diagram, Bar diagram, Histogram, Line plot, frequency curves & polygons, Scatter Plot, Gantt Chart, Box Plot.
3. **Descriptive Statistics:** Measures of Central Tendencies, Dispersions, Relative measures of Dispersions, Moments, Skewness, Kurtosis.
4. **Parametric Tests**: Testing for Mean(s), Variance(s), Proportion(s), ANOVA for one-way two-way and two-way with one and m-observations per cell and with & without interactions,
5. **Non–Parametric tests:** Sign test, Wilcoxon Sign Rank test, Mann-Whitney U-test, Run test, Kolmogorov Smirnov test, Chi-square test for goodness of fit and Chi-square test independence.
6. **Design and Analysis of Experiments:** Analysis of Variances for Completely randomized, randomized block and Latin Square Designs.
7. **Regression Analysis:** Analysis of Simple and Multiple Linear Regression models, Selection Best Linear Regression Model (All possible, forward, backward, stepwise and stage wise methods). Binary and multinomial Logistic regression models, Probit analysis.
8. **Multivariate Data Analysis:** Linear Discriminant Analysis, Principal Component analysis, Factor analysis, multi-dimensional scaling, Cluster analysis.

M.Sc. (Data Science) III-Semester Scheme

Paper Code	Title of the paper	Credits	No of Hours per week	No of Hours Exam duration	Maximum Internal marks	Semester End exam marks
Theory Papers						
MDS-301	Deep Learning Techniques	3	3	3	30	70
MDS-302	Computer Networks	3	3	3	30	70
MDS-303 (E-I)	a) Text Data Analytics b) Statistical Pattern Recognition c) Cloud Computing	3	3	2	30	70
MDS-304 (E-II)	a) Big Data Analytics b) Image Processing & Analysis c) Web Technologies	3	3	2	30	70
Practical Papers						
MDS-305	Deep Learning & Computer Networks Lab using Python	2	4	2	-	50
MDS-306	E-I & E-II using Python	2	4	2	-	50
MDS-307	Mini Project	4	8	-	-	100
	No of Credits	**20**	**12+16**		**120**	**480**

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-301: PAPER- I: DEEP LEARNING TECHNIQUES

UNIT – I

Artificial Neural Networks: Introduction, Biological Activations of Neuron; Artificial Neuron Models: McCulloch-Pitts, Perceptron, Adaline, Hebbian Models; Characteristics of ANN, Types of Neuron Activation Function, Signal functions and their properties, monotonicity, ANN Architecture, Classification Taxonomy of ANN, Supervised, Un-supervised and Reinforcement learning; Learning tasks, Memory, Adaptation, Statistical nature of the learning process. Statistical learning theory**.** Gathering and partitioning of data for ANN and its pre and post processing.

UNIT – II

Supervised learning algorithms: Perceptron Learning Algorithm, Derivation, Perceptron convergence theorem (statement); Multi-layer Perceptron Learning rule, limitations. Applications of the Perceptron learning. Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Feed-forward and Feed-back Back-Propagation Algorithms and derivation.

Unsupervised learning Algorithms: Hebbian Learning, Competitive learning**.** Self-Organizing Maps, SOM algorithm, properties of feature map, computer simulations, Vector quantization, Learning vector quantization.

UNIT – III

Radial Basis Function Networks, Approximation properties of Radial Basis Function Networks. Boltzmann Machine, Hopfield model. Reinforcement learning, Markov Decision Process, Hidden Markov Model, Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory Networks, Generative Adversarial Networks, Deep belief Networks.

REFERENCES:

1. Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York: Macmillan Publishing. A comprehensive book and contains a great deal of background theory
2. Yagnanarayana, B. (1999): “Artificial Neural Networks” PHI
3. Bart Kosko(1997): Neural Networks and Fuzzy systems, PHI
4. Jacek M. Zurada(1992): Artificial Neural Systems, West Publishing Company.
5. Carling, A. (1992). Introducing Neural Networks. Wilmslow, UK: Sigma Press.
6. Box and Jenkins: Time Series analysis, Springer

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-302: PAPER- II: COMPUTER NETWORKS

UNIT – I

Computer Networks Fundamentals: Overview, Network Hardware, Network Software, Reference models– OSI Model, TCP/IP Reference Model, Comparison of OSI and TCP/IP Reference Model, Example Networks, Network Standardization. Physical Layer: Guided Transmission Media, Wireless Transmission, Multiplexing, Switching. Data Link Layer: Design Issues, Error Detection and Correction, Data Link Layer Protocols, Sliding Window Protocol

UNIT – II

Multiple Access Sublayer: ALOHA, CSMA, Collision Free Protocols, Ethernet, Wireless LAN802.11, Data Link Layer Switching –Repeaters, Hubs, Bridges, Switches, Routers, Gateways. Network Layer: Design Issues, Routing Algorithms – Shortest path, Flooding, Distance Vector Routing, Link state Routing, Hierarchical, Broadcast Routing, Multicast Routing; Congestion Control Algorithms. Internetworking: Tunneling, Internetwork Routing, Fragmentation, IPv4 Vs IPv6Protocol, IP Addresses, CIDR, Internet Control Protocols–IMCP, ARP, RARP, DHCP.

UNIT – III

Transport Layer: Services provided to the upper layers, Transport Protocols, Overview of Congestion Control. The Internet Transport Protocols: Introduction to UDP&RPC, Real Time Transport Protocols, The Internet Transport Protocols–TCP, TCP Service Model, TCP protocol, TCP Segment Header, TCP Connection Establishment, TCP Connection Release, Modelling TCP Connection Management, TCP Sliding Window, TCP Time Management, TCP Congestion Control. Application Layer: DNS, TELNET, E-Mail, FTP, HTTP, SSH, Overview of WWW.

REFERENCES

1. Andrew S. Tanenbaum, David J Wetherall, Computer Networks (5th edition)
2. William Stallings, Data and Computer Communications
3. Behrouz A. Forouzan, Data Communication and Networking
4. Behrouz A Forouzan, Firouz Mosharraf, Computer Networks A Top-Down Approach

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-303 (A): PAPER- III (A): TEXT DATA ANALYTICS

UNIT - I

Introduction to Natural Language Processing Basic, Language Syntax and Structure (Words, Phrases, Clauses, & Grammar), Language Semantics Processing, (Lexical Semantic Relations, Homonyms, Homographs, and Homophones, Capitonyms, Hyponyms and Hypernyms), Text Corpora (Corpora Annotation and Utilities), Accessing Text Corpora (Brown Corpus, WordNet Corpus) and NLP Applications (Machine Translation, Text Summarization and Text categorization).

UNIT – II

Concept of the Tokenization, Sentence Tokenization, Word Tokenization, Concept of the Text Normalization, (Cleaning Text, Removing Special characters, removing stop words, etc) correcting words using stemming and Lemmatization and Understanding text syntax and structure. (POS tagging and Parsing)

UNIT – III

Concepts of feature extraction, Methods of Feature extraction (Bag of words Model, TF-IDF Models, Advanced word Factorization Models likes Word2vec), Strengths and weakness of models and Word cloud … etc, Concepts of Document term matrix, Term Document Matrix.

Concepts of Topic Modelling, Algorithms of Topic Modelling (Latent Semantic Indexing (LSI), Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF) and Similarity based text clustering models), Text Classification using supervised methods (Like Multinomial Naïve Bayes, Support vector machines, Random Forest etc.), concept of Sentiment Analysis and its applications.

REFERENCES

1. Chapman & Hall: Handbook of Natural Language Processing, Second Edition.
2. CRC: Machine Learning & Pattern Recognition, 2nd Edition.
3. Christopher Manning and Hinrich Schuetze: Foundations of Statistical Natural Language Processing.
4. Dipanjan Sarkar: Text Analytics with Python, A press Publication.
5. Julia Silge: Text Mining with R: A Tidy Approach, 1st Edition.

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-303 (B): PAPER- III (B): STATISTICAL PATTERN RECOGNITION

UNIT - I

Basic Concepts to Statistical Pattern Recognition, Pattern Recognition System, Fundamental problems in Pattern Recognition. Linear classifiers: Linear Discriminant Function (for binary outputs) with minimum squared error; Linear Discriminant function (for the normal density), Error bounds for Normal density. Statistical Decision Theory: Introduction, Bayes theorem, Bayes Decision Theory (continuous and discrete features), Bayes Classifier. Simple problems.

UNIT – II

Probability of errors: Two classes, Normal distribution, equal covariance matrix assumptions, Chernoff bounds and Bhattacharya distance. Nearest Neighbour Decision rules: Nearest Neighbor Algorithm for classification, K-Nearest Neighbor Estimation. Variants of the Nearest Neighbor Algorithm, description convergence, finite sample considerations. Estimation of probability of error in case Nearest Neighbour and Bayes classifiers. Minimum Error Rate Classifier, Estimation of Probabilities. Comparison of Nearest Neighbour with the Bayes Classifier. Simple problems

UNIT – III

Hidden Markov Model and its use for pattern recognition. Branch and Bound Technique for the use of classification. Neural Networks: Perception linear classifier. Support Vector Machines: construction of Support Vectors, Support Vector Machines algorithm for Classification. Simple problems. Combination of Classifiers: Introduction, Methods for Constructing Ensembles of Classifiers, Methods for Combining Classifiers.

Feature selection and extraction: Feature extraction and Feature selection techniques Inter and intra class distance measures, Probabilistic distance measures, Principal Components Analysis for variable selection and dimensionality reduction.

REFERENCES

1. R.O. Duda & H.E. Hart(1978): Pattern Recognition and scene analysis, Wiley
2. Earl Gose, Richard Johnson Baugh and Steve Jost (2009): Pattern Recognition and Image Analysis, PHI.
3. Pattern Recognition: An Algorithmic Approach: Murty, M. Narasimha, Devi, V. Susheela, Spinger Pub,1st Ed.
4. Duda, Hast & Strok: Pattern Recognition.
5. Pattern Recognition: An Algorithmic Approach: Murty, M. Narasimha, Devi, V. Susheela, Spinger Pub,1st Ed.

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-303 (C): PAPER- III (C): CLOUD COMPUTING

UNIT- I

Introduction, Benefits and challenges, Cloud computing services, Resource Virtualization, Resource pooling sharing and provisioning, Case study of Iaas, Paas and Saas

Scaling in the Cloud, Capacity Planning, Load Balancing, File System and Storage, Containers. Multi-tenant Software, Data in Cloud, Database Technology.

UNIT-II

Content Delivery Network, Security Reference Model, Security Issues, Privacy and Compliance Issues**,** Portability and Interoperability Issues, Cloud Management and a Programming Model Case Study, Popular Cloud Services.

UNIT- III

Enterprise architecture and SOA, Enterprise Software, Enterprise Custom Applications, Workflow and Business Processes, Enterprise Analytics and Search, Enterprise Cloud Computing Ecosystem.

REFERENCE BOOKS

1. Sandeep Bhowmik (2017): Cloud Computing, Cambridge University Press.
2. Gautam Shroff (2016): Enterprise Cloud Computing - Technology, Architecture, Applications by Cambridge University Press.
3. Kai Hwang, Geoffrey C. Fox, Jack J. Dongarra (2012): Distributed and Cloud Computing From Parallel Processing to the Internet of Things, Elsevier.

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-304 (A): PAPER- IV(A): BIG DATA ANALYTICS

UNIT-I

Foundations of Big Data & The Hadoop Ecosystem: Introduction to Big Data, Elements of Big Data Analytics (Volume, Velocity, Variety, Veracity). Business Contexts & Use Cases: Social Media Analytics, Fraud Detection (General and in Insurance), Retail and Customer Analytics

Technologies for Handling Big Data: Core Principle: Distributed and Parallel Computing. Introducing Hadoop: The solution to big data challenges. Hadoop Ecosystem: Storage: HDFS (Hadoop Distributed File System) Processing: MapReduce Framework Resource Management: Hadoop YARN (Introduction) Data Access & Ingestion Tools (Overview): HBase (NoSQL Database) Sqoop (Data transfer between Hadoop and RDBMS) Flume (Ingesting log/streaming data)

UNIT-II

Data Processing with MapReduce & Data Storage: Deep Dive into MapReduce: The MapReduce Framework (Mapper, Shuffle & Sort, Reducer). Techniques to Optimize MapReduce Jobs (Combiners, Partitioners). Developing a Simple MapReduce Application.

Data Storage Paradigms: RDBMS vs. Big Data, Limitations of traditional databases. Introduction to NoSQL Databases: HBase and its role in Big Data processing. Integrating Big Data with Traditional Data Warehouses. 2.3. Customizing & Controlling MapReduce: Controlling Execution with Input/Output Formats. Optimizing jobs with Combiners. Implementing a MapReduce program for a classic problem (e.g., Sorting).

UNIT – III

Data Analysis with High-Level Tools (Hive, Pig) & Modern Architecture: Modern Hadoop Architecture: Hadoop YARN: Architecture, Advantages, and Working. Data Warehousing & SQL-on-Hadoop: Exploring Hive: Introduction & Hive Services, Hive Data Types, DDL, and DML , Data Retrieval Queries and JOINS in Hive, Data Flow Programming: Analyzing Data with Pig: Introducing Pig and Pig Latin Working with key Operators and Functions Running and debugging Pig Scripts. Introduction to Analytics: Understanding Analytics vs. Reporting. Types of Analytics (Descriptive, Diagnostic, Predictive, Prescriptive).Overview of Analytical Approaches and Tools.

REFERENCES:

1. DT Editorial Services, Big Data – Black Book (dream tech)
2. Radha S, M. Vijayalakshmi, Big Data Analytics
3. Arshdeep B and Vijay M, Big Data Science & Analytics – A Hands-On Approach.
4. Frank Ohlhorst, Big Data Fundamentals – Concepts, Drivers, Techniques

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-304 (B): PAPER- IV(B): IMAGE PROCESSING AND ANALYSIS

UNIT – I

Digital Image Fundamentals & Image Transforms: Digital Image Fundamentals, Sampling and Quantization, Relationship between Pixels, Image Transforms: 2-D FFT, Properties, Walsh Transform, Hadamard Transform, Discrete Cosine Transform, Haar Transform, Slant Transform, Hoteling Transform. Image Enhancement (Spatial Domain): Introduction, Image Enhancement in Spatial Domain, Enhancement Through Point Operation, Types of Point Operation, Histogram Manipulation,

UNIT -II

Linear and Non-Linear Gray Level Transformations, Local or Neighbourhood Criterion, Median Filter, Spatial Domain High-Pass Filtering. Image Enhancement (Frequency Domain): Filtering in Frequency Domain, Low Pass (Smoothing) and High Pass (sharpening) Filter in Frequency Domain. **Image Restoration:** Degradation Model, Algebraic Approach to Restoration, Inverse Filtering, Least Mean Square Filters, Constrained Least Squares Restoration, Interactive Restoration.

UNIT -III

Image Segmentation: Detection of Discontinuities, Edge Linking And Boundary Detection, Thresholding, Region Oriented Segmentation. Morphological Image Processing: Dilation and Erosion: Dilation, Structuring Element Decomposition, Erosion, Combining Dilation and Erosion, Opening and Closing, The Hit or Miss Transformation

Image Compression: Redundancies and their Removal Methods, Fidelity Criteria, Image Compression Models, Huffman and Arithmetic Coding, Error Free Compression, Lossy Compression, Lossy and Lossless Predictive Coding, Transform Based Compression, JPEG 2000 Standards

REFERENCES:

1. Rafael C. Gonzalez, Richard E. Woods: Digital Image Processing –, 3rd Edition, Pearson, 2008.
2. S Jayaraman, S Esakkirajan, T Veerakumar: Digital Image Processing- - TMH, 2010.
3. Scotte Umbaugh: Digital Image Processing and Analysis-Human and Computer Vision Application with using CVIP Tools, 2nd Ed, CRC Press, 2011
4. A.K. Jain: Fundamentals of Digital Image Processing, PHI, 1989
5. Somka, Hlavac, Boyle: Digital Image Processing and Computer Vision, Cengage Learning (Indian edition) 2008.
6. John C. Russ, J. Christian Russ: Introduction to Image Processing & Analysis, CRC Press, 2010.

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-304 (C): PAPER- IV (C): WEB TECHNOLOGIES

UNIT-I

HTML: Markup languages, common tags, header, test styling, linking images Formatting text, Unordered lists, nested and ordered list, Tabs-and formatting, Basic forms; Complex forms linking, Meta Tags.

Dynamic HTML: Cascading style sheets in line styles, style element, External Style sheet, text flow and Box model, user style sheets.

UNIT-II

Introduction to scripting: Java Script, Data types, Arithmetic's Equality relational, assignment increment, decrement operators, Java Script Control Structures- if, if-else, while. Java Script Control Structures: For, Switch, Do/while, break.

Event Model: Onclick, Onload, Error Handling, On Errors On mouse move, On mouse over, On mouse out, On focus, On blur, On submit.

Programming modules, recursion, recursion vs iteration global functions arrays, using arrays, Reference and reference parameters, passing arrays to functions, multiple subscripted arrays, objects-math, string. Boolean and number.

UNIT-III

Web Servers: Personal Web server, Internet information server, Apache Web Server, Installation of a Web Server. Active Server Pages, Client-side Scripting vs Server-side Scripting, Server-side Active X Component, ADO, file system objects, Session tracking, CGI and PERL5, String Processing and Regular Expressions, Server side includes, Cookies and PERL XML Document Type Definition, XML Parsers, Using XML with HTML.

REFERENCES:

1. Deiterl, Deitel & Nieto: "Internet & World Wide Web - How to Program", Pearson Education, 3e, 2004.
2. Steven Holzner: "HTML Black Book - Comprehensive Problem Server", Dream Tech Press, 2000.
3. B Sosinsky, V Hilley: "Programming the Web - An Introduction", MGH, 2004.

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-305 (LAB-I): PAPER- V:
DEEP LEARNING & COMPUTER NETWORKS USING PYTHON (LAB)

List of Experiments on Deep Learning

1. Implementation of Perceptron and Multi-layer Perceptron Learning Algorithm.
2. Implementation of Gradient Descent Learning rule
3. Implementation of Least Mean Square learning law
4. Implementation of Widrow-Hoff Learning rule
5. Implementation of Back-Propagation learning Algorithms.
6. Implementation of Hebbian Learning,
7. Implementation of Competitive learning
8. Implementation of Markov Decision Process,
9. Implementation of Hidden Markov Model,
10. Implementation of Convolutional Neural Networks,
11. Implementation of Recurrent Neural Networks,
12. Implementation of Long-Short Term Memory.

List of Experiments on Computer Networks:

1. TCP Echo Server & Client: Write a Python program where a client sends a message to a server, and the server echoes it back.
2. UDP Time Server: Write a program where a UDP client requests the current time from a server, and the server responds with its system time.
3. Simulate a sender and receiver that implement the Stop-and-Wait protocol with ACK/NACK for error control. Introduce a random probability for packet loss.
4. Implement a simple version of the Go-Back-N sliding window protocol to demonstrate flow control.
5. Write a Python script using scapy to Capture ICMP packets (from a ping command).
   Parse and print the fields of the captured IP and ICMP headers (e.g., Version, TTL, Type, Code).
6. Write a program to simulate the functionality of nslookup or dig.
7. Develop a simple console-based chat application.
   a) Create a multi-threaded server that can accept connections from multiple clients and broadcast messages from one client to all others.
   b) Create a peer-to-peer chat where users can send messages to each other's IP addresses and ports.

M.SC. (DATASCIENCE) III-SEMESTER SYLLABUS

MDS-306 (LAB-II): PAPER- VI: E-I & E-II

Elective-I:

E-I: A) TEXT DATA ANALYTICS

1. Performing text data understanding and Pre-processing steps with Python using Scikit-learn for the data sets. (Data transformation: Standardization, normalization, encoding, discretization, imputation of missing values).
2. Perform data collection by web scrapping using python. (It includes the tasks of identifying the URL that you want to scrape, Inspecting the Page the information required to gather, find the data you want to extract, writing the code, Run the code and extract the data and store the data in the required format.)
3. Text Classification algorithms implementation
   (i) Logistic Regression
   (ii) Naïve Bayes classifier
   (iii) Hidden Markov model
   (iv) Decision Tree algorithm (ID3)
4. Text clustering algorithms implementation.
   (i) K-means algorithm
   (ii) KNN
5. Principal Component and Linear Discriminant analysis
6. Dataset Description: Tasks to be performed:
   (i) How to decide variance threshold in data reduction? Code Snapshot, Output Snapshot,
   (ii) Perform following Data Pre-processing tasks using python, Code Snapshot, Output Snapshot.
   (iii) Data reduction using variance threshold, univariate feature selection, recursive feature elimination, PCA, correlation

E-I: B) STATISTICAL PATTERN RECOGNITION

1. Computation of Linear discriminant classifier function for two-multivariate normal classes.
2. Computation of Linear discriminant classifier function using Minimum Squared function for Binary data.
3. Bayes classifier and the computation of its Error rate (Probability of Error).
4. Nearest Neighbor Classifier and computation of its error rate.
5. Classification using Hidden Markov Model.
6. Feature selection using Principal Component Analysis.

E-I: C) CLOUD COMPUTING

1. Install Virtual box / VMware Workstation with different flavours of Linux or windows OS on top of windows7 or 8.
2. Install a C compiler in the virtual machine created using virtual box and execute Simple Programs.
3. Install Google App Engine. Create hello world app and other simple web applications using python/java.
4. Use GAE launcher to launch the web applications.
5. Simulate a cloud scenario using Cloud Sim and run a scheduling algorithm that is not present in Cloud Sim.
6. Find a procedure to transfer the files from one virtual machine to another virtual machine.
7. Find a procedure to launch virtual machine using try stack (Online Open stack Demo Version)
8. Install Hadoop single node cluster and run simple applications like wordcount.

PRACTICAL LISTS OF ELECTIVE-II:

E-II: A) BIG DATA ANALYTICS

1. Installation of Hadoop Framework, it‘s components and study the HADOOP ecosystem.
2. Write a program to implement word count program using MapReduce
3. Experiment on Hadoop Map-Reduce / PySpark: -Implementing simple algorithms in Map-Reduce: Matrix multiplication.
4. Install and configure MongoDB/ Cassandra/ HBase/ Hypertable to execute NoSQL Commands.
5. Implementing DGIM algorithm using any Programming Language/ Implement Bloom Filter using any programming language
6. Implement and Perform Streaming Data Analysis using flume for data capture, PYSpark / HIVE for data analysis of twitter data, chat data, weblog analysis etc.
7. Implement any one Clustering algorithm (K-Means/CURE) using Map-Reduce.
8. Implement Page Rank Algorithm using Map-Reduce

E-II: B) IMAGE PROCESSING & ANALYSIS

1. Display of Gray scale images
2. Histogram Equalization
3. Design non-linear filtering
4. Determination of edge detection using operators
5. 2-D DFT and DCT
6. Filtering in Frequency domain
7. Display of colour images
8. Conversion between colour spaces
9. DWT of images
10. Segmentation using watershed transform

E-II: C) WEB TECHNOLOGIES

1. Develop College Website using XHTML and CSS.
2. Develop HTML form with client validations using Java Script.
3. Publishing XML document using XSLT
4. XML document processing using SAX and DOM.
5. Text processing using Regular expressions and pattern matching.
6. Develop form processing application using CGI.pm
7. Develop CGI-Perl Web application with State and Session Tracking.
8. Develop a simple Java servlet application.
9. Develop Java servlet application with session tracking
10. Develop a simple JSP application.
11. Creation of an application access database with JDBC
12. Develop full-fledged web application with database access spreading over to 3 sessions.

M.SC. (DATASCIENCE) III-

SEMESTER MDS-307 (LAB-III): PAPER-

VII: MINI PROJECT

PROJECT GUIDELINES:

1. The Head of the Department will appoint Internal supervisor to guide the students for each group.
2. Each group should consist of 3-5 students only (exempted for industry internship students).
3. Each group has to search for the internship from any industry/ institution, if not found they have to choose a project with the help of supervisor allotted such that, the aim of project work is to develop solutions to realistic problems applying the knowledge and skills obtained on the courses studied with specializations, new technologies and current industry practices.
4. Each student in the group must actively participate and report to the internal supervisor in each week.
5. Each student has to give minimum two seminars, one in the second week ("Project Design Seminar") another on 8^{th} week (project progress seminar).
6. Submit Title of the project and one page abstract /synopsis about the project in the first week to the Head, forwarded by the internal supervisor.
7. Each project should give a 30 minutes presentation using power point presentation and followed by 10 minutes of discussion.

8. Project seminar presentations should contain, source of the data, Sample data, data description, literature survey on the similar studies, objectives of the study, Methodology, statistical techniques, work plan etc. and details of progress of the work, individual roles and their work distribution and their plan etc.
9. Each group Project Report should follow the Ph.D. thesis norms with Plagiarism report and each group has to submit two copies duly signed by the Students, Supervisor, industry certificate (if exists) and Head of the Department on before the last instruct date of the semester.
10. Project will be awarded based on all stages of the project and the topic chosen, seminar presentation, communication skills, role/ contribution of the student in the project etc and viva-voce conducted by the internal & External examiners.
11. The Project dissertation should follow Ph.D. typing format: Font: Times New Roman 12 size; Titles of Capital Letters, Sub titles bold small letters, Double spacing; Paragraph indentation, justify alignment, Numbering should follow as per norms (2chapter.3subheading.4section: 2.3.4) all equations also should be numbered as per norms. numbering for tables on top with heading and numbering for fig in bottom, etc.
12. The Data Science Project Report submitted to be hard bound (one for exam branch and one for Institution) should contains:
    1. **Title Page** in prescribed format
    2. **Declaration** by the student in prescribed format
    3. **Certificate** from the Industry / supervisor in prescribed format
    4. **Acknowledgments**
    5. **Contents** in prescribed format
    6. The contents of project dissertation should contains/ covers the chapters are:
        a. **Introduction** (Introduction, Motivation to topic, Significance and need of the study, Problem Statement, Objectives of the study, Chapter wise summary).
        b. **Review of the Literature** (Introduction with list of authors contributed (collected from journals) related to topic, Each author method description (5-10 methods) and its comparison study
        c. **Materials and Methods** (sample data set, Sample description, Software's used, Tools & Technique applied)
        d. **Program Code for Implementation**
        e. **Exploratory Data Analysis** (Outputs of descriptive statistics, Data visualizations, Data interpretations and conclusions)
        f. **Model Building** (Out puts of the Models used for the data sets, Data interpretations & conclusions)
        g. **Conclusions & Future Scope**
        h. **Appendix** (if any)
        i. **Bibliography** (Minimum 30 articles)
        j. **Plagiarism Report**

M.Sc. (Data Science) II-Year, IV-Semester

Paper Code	Title of the paper	Credits	No of Hours per week	No of Hours Exam duration	Maximum Internal marks	Semester End exam marks
Theory papers						
MDS-401	Cryptography	3	3	2	30	70
MDS-402	Data Mining	3	3	2	30	70
MDS-403 E-III	a) Sentimental Analysis b) Computer Vision c) Machine Learning Operations	3	3	2	30	70
MDS-404 E-IV	a) Data stream mining b) Web mining c) Scalable Architecture	3	3	2	30	70
Practical papers						
MDS-405	Cryptography & Data Mining	2	4	2	-	50
MDS-406	E-III & E-IV	2	4	2	-	50
	Capstone Project	4	8	-	-	100
	No of Credits	**20**	**12+16**		**120**	**480**

M.SC. (DATA SCIENCE) IV-SEMESTER SYLLABUS
MDS-401: PAPER- I: CRYPTOGRAPHY

UNIT – I

Overview of Network Security: OSI Security Architecture, Security Attacks, Security Services, Security Mechanisms, a Model for Network Security. **Classical Encryption Techniques**: Symmetric Cipher Model, Substitution Techniques, Transposition Techniques, Rotor Machines, Steganography. **Block Ciphers** : Structure and Data Encryption Standard (DES),Strength of DES. **Block Cipher Operation**: Double and Triple DES, Electronic Code Book, Cipher Block Chaining Mode, Cipher Feedback Mode, Output Feedback Mode, Counter Mode.

UNIT – II

Encryption Standard (AES): Origins, Structure, Round Functions, AES Key Expansion. **Pseudorandom Number Generation and Stream Ciphers**: Principles, Block Cipher based PRNG, RC4. **Public-Key Cryptography**: Principles of Public-Key Cryptosystems, RSA Algorithm. K**ey Management and Distribution**: Symmetric and Asymmetric Key Distribution, Public Key Distribution, X.509 Certificates, Diffie-Hellman Key Exchange.

UNIT – III

Cryptographic Hash Functions: Applications, SHA & MD5 Algorithm. **Message Authentication Codes (MAC):** requirements, HMAC, CMAC. **Digital Signatures**: Concepts, NIST Digital Signatures Algorithm. **Transport-Level Security**: SSL, TLS, HTTPS, SSH. E-Mail Security: Pretty Good Privacy, S/MIME. **IP Security**: Overview, Architecture, Encapsulating Security Payload, Internet Key Exchange. **System Security:** Intruders, Intrusion Detection, Password Management, Virus and Countermeasures, Firewall Design Principles and Types.

REFERENCES:

1. William Stallings, Cryptography and Network Security – Principles and Practice (6e)
2. Zhenfu Cao, New Directions of Modern Cryptography
3. Douglas R. Stinson, Cryptography Theory and Practices
4. Tom St Denis, Simon Johnson, Cryptography for Developers
5. Joseph Migga Kizza, A Guide to Computer Network Security
6. A. Menezes, P. Van Oorschot, S. Vanstone, Handbook of Applied Cryptography
7. Henk C.A. van Tilborg, Sushil Jajodia, Encyclopedia of Cryptography and Security
8. Keith M. Martin, Everyday Cryptography–Fundamental Principles and Applications

M.SC. (DATASCIENCE) IV-SEMESTER SYLLABUS
MDS-402: PAPER- II: DATA MINING

UNIT-I

Introduction to Data Mining and Data Understanding: Data Mining Concepts – Definition, Need for Data Mining, **Data Mining Scope** – Types of Data to be Mined, Types of Patterns to be Mined, **Technologies for Data Mining** – Supporting Tools and Techniques, **Applications of Data Mining** – Targeted Domains and Use Cases, **Major Issues in Data Mining** – Challenges and Research Directions, **Getting to Know Your Data-** Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization Techniques, Measuring Data Similarity and Dissimilarity.

UNIT-II

Frequent Pattern Mining and Classification: Frequent Pattern Mining & Associations- Basic Concepts and Methods, Frequent Itemset Mining Techniques, Interestingness of Patterns, Pattern Evaluation Methods, **Classification: Basic Methods -**Concepts of Classification, Decision Tree Induction, Bayes Classification Methods. **Classification: Advanced Methods-** Bayesian Belief Networks, Classification by Backpropagation, Support Vector Machines (SVM)

UNIT-III

Cluster Analysis and Data Mining Trends: Cluster Analysis: Concepts and Methods- Introduction to Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Evaluation of Clustering. **Data Mining Trends and Research Frontiers-** Mining Complex Data Types, Alternative Methodologies in Data Mining, Applications of Data Mining, Data Mining and Society, Emerging Trends in Data Mining

REFERENCES:

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, lst Edition,2009.
3. Pang-Ning Tan, Michael Steinbach, Vipin kumar, Introduction to Data Mining, pearcon Education,2008.

M.SC. (DATASCIENCE) IV-SEMESTER SYLLABUS
MDS-403 A: PAPER- III (A) : SENTIMENTAL ANALYSIS

UNIT-I

Basics & Applications: Introduction, Applications, Research Scope, Sentiment Analysis as Mini NLP. **The problem of Sentiment Analysis**: Definition & Opinion Summary - Affect, Emotion, and Mood. Deferent types of opinions, Author Vs. Reader standpoint.

Document-Level sentiment classification: Supervised and Unsupervised Sentiment classification, Sentiment Rating Prediction, Cross-Domain and Cross-Language Sentiment Classification, Emotion Classification of Documents.

UNIT-II

Subjectivity, Sentence-Level Analysis & Lexicons: Subjectivity & Sentence Sentiment Classification – Sentence Subjectivity, Sentiment Classification, Handling Conditional & Sarcastic Sentences, Cross-Language Classification, Discourse-Based Sentiment, Emotion Classification of Sentences. **Sentiment Lexicon Generation** – Dictionary-Based Approach, Corpus-Based Approach, Desirable vs. Undesirable Facts.

UNIT-III

Comparative Opinions, Summarization & Opinion Quality: Analysis of Comparative Opinions – Problem Definition, Identifying Comparative Sentences, Preferred Entity Set, Types of Comparison, Entity & Aspect Extraction. **Opinion Summarization & Search** – Aspect-Based Summarization, Contrastive View, Traditional Summarization, Summarization of Comparative Opinions, Opinion Search & Retrieval Techniques. **Mining Intentions** – Intention Mining Problem, Intention Classification, Fine-Grained Mining. **Fake & Low-Quality Opinions** – Fake/Deceptive Opinion Detection (Spam Types, Supervised Detection, Behavioral Analysis, Group Spam, Multiple IDs, Business Exploitation), Quality of Reviews (Regression Approach & Other Methods).

REFERENCES

l. Bing Liu "Sentiment Analysis University Press, 2015. Mining Opinions, Sentiments and Emotions, Cambridge

M.SC. (DATASCIENCE) IV-SEMESTER SYLLABUS
MDS-403 B: PAPER-III (B): COMPUTER VISION

UNIT-I

Computer vision Introduction: Computer vision-Image Formation: Geometric primitives and transformation-Photometric image formation-The digital camera.

UNIT-II

Image Processing: Point Operation-Linear filtering-More neighbourhood operators, Fourier Transforms-Pyramids and wavelets - Geometric Transformations - Global optimization.

UNIT-III

Feature Detection and Segmentation: Feature Detection & Matching – Points and Patches, Edges, Lines. **Segmentation** – Active Contours, Split & Merge, Mean Shift & Mode Finding, Normalized Cuts, Graph Cuts & Energy-Based Methods. **Recognition:** Object Detection, Face Recognition, Instance Recognition, Category Recognition, Context & Scene Understanding, Recognition Datasets and Test Sets.

REFERENCES

1. Richard Szeliski (201l): Computer Vision-Algorithms and Applications, Springer verlang London Limited.
2. Deep Leaming, by Goodfellow, Bengio, and Courville.
3. Dictionary of Computer Vision and Image processing, by Fisher et al.

M.SC. (DATASCIENCE) IV-SEMESTER SYLLABUS
MDS-403 C: PAPER- III (C): MACHINE LEARNING OPERATIONS

UNIT – I

Introduction to ML Ops & Data Management: Overview and Significance of ML Ops in Industry, Challenges in Deploying and Managing ML Models, Principles of DevOps Applied to ML Ops, Cloud Platforms and Tools for ML Ops, Data Versioning and Management in ML Projects, Reproducibility, Data Lineage, and Version Control (Git), Collaborative Development Workflows

UNIT – II

Model Deployment and Monitoring: Model Packaging and Containerization, Deploying Models to Production Environments. Infrastructure Orchestration and Scaling for ML Workloads, Monitoring Model Performance and Handling Concept/Data Drift

UNIT – III

Continuous Integration & Delivery (CI/CD) for ML Ops: Automated Testing and Validation of ML Models, Continuous Integration Pipelines for ML Ops, Continuous Deployment and Release Strategies, Feedback Loops and Continuous Improvement

REFERENCES

1. Building Machine Learning Pipelines by Hannes Hapke and Catherine Nelson
2. Data Science on AWS: Implementing End-to-End, Continuous AI and Machine Learning Pipelines by Chris Fregly and Antje Barth
3. MLOps with Azure: Machine Learning Operations on Azure by Srinivasa Rao and Amit Kumar Saha

M.SC. (DATASCIENCE) IV-SEMESTER SYLLABUS
MDS-404 A: PAPER- IV (A) : DATA STREAM MINING

UNIT-I

Stream Data Mining and Evaluation: MOA (Massive Online Analysis) Stream Mining – Concepts and Assumptions. **Requirements and Mining Strategies** – Key Requirements, Strategies for Stream Mining. **Change Detection Strategies** – Handling Concept Drift in Data Streams. **MOA Experimental Settings** – Framework and Environments. **Evaluation of Data Streams** – Previous Practices, Procedures, and Testing Frameworks. **Stream Data Characteristics** – Data Sources, Generation Speed, Data Size, Evolving Stream Settings.

UNIT-II

Hoeffding Trees and Evolving Data Streams: Hoeffding Trees – Concept, Hoeffding Bound for Tree Induction, Basic Algorithm. **Enhancements** – Memory Management, Numeric Attributes, Batch vs. Data Stream Approaches. **Prediction Strategies** – Majority Class, Naïve Bayes Leaves, Adaptive Hybrid. **Hoeffding Tree Ensembles** – Data Stream Settings, Realistic Ensemble Sizes. **Evolving Data Streams** – Algorithms for Mining with Change, Adaptive Stream Mining Methodology. **Change Detection** – Optimal Change Detector and Predictor, Adaptive Sliding Windows, Maintaining Updated Windows of Varying Length

UNIT-III

Adaptive Hoeffding Trees and Ensemble Methods: Adaptive Hoeffding Trees – Introduction and Concept. Decision Trees on Sliding Windows. Hoeffding Adaptive Trees. **Adaptive Ensemble Methods** – Bagging with Trees of Different Sizes, Bagging with ADWIN. Adaptive Hoeffding Option Trees. Performance Evaluation of Methods

REFERENCES:

1. Data Stream Mining: A Practical Approach by Albert Bifet and Richard Kirkby.
2. Knowledge discovery from data streams by Gama
3. Machine Learning for Data Streams by Albert Bifet, Ricard Gavalda; MIT Press,2017.

M.SC. (DATASCIENCE) IV-SEMESTER SYLLABUS
MDS-404 B: PAPER- IV (B) : WEB MINING

UNIT – I

Web Data Mining & Data Mining Foundations: Introduction to WWW, Web Mining, and Data Mining, Association Rule Mining – Apriori Algorithm, Frequent Itemset & Rule Generation, Multiple Minimum Supports, Class Association Rules. Sequential Pattern Mining – GSP, PrefixSpan, Rule Generation from Patterns

UNIT – II

Machine Learning for Web Mining: Supervised Learning – Decision Trees, Rule Induction, Classification based on Associations, Naïve Bayes & Text Classification. **Unsupervised Learning** – K-means Clustering, Hierarchical Clustering (Single Link, Complete Link, Average Link), Strengths & Weaknesses

.

UNIT – III

Information Retrieval, Link Analysis & Web Crawling: Information Retrieval – Boolean Model, Vector Space Model, Statistical Language Model, Relevance Feedback, Evaluation Measures. Text & Web Page Preprocessing – Stopword Removal, Stemming, Duplicate Detection, Inverted Index & Compression, Latent Semantic Indexing. Web Search & Issues – Web Search, Meta Search, Web Spamming. Link Analysis – PageRank, HITS, Community Discovery. Web Crawling – Crawler Algorithms (BFS, Focused, Topical), Implementation Issues, Ethics. Sentiment Classification – Sentiment Phrases, Text Classification Methods.

REFERENCES :

1. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data by Bing Liu (Springer Publications)
2. Data Mining: Concepts and Techniques, Second Edition Jiawei Han, Micheline Kamber (Elsevier Publications)
3. Web Mining:: Applications and Techniques by Anthony Scime
4. Mining the Web: Discovering Knowledge from Hypertext Data by Soumen Chakrabarti

M.SC. (DATASCIENCE) IV-SEMESTER SYLLABUS
MDS-404 C: PAPER- IV(C): SCALABLE ARCHITECTURE

UNIT-I

Scalable Applications and Big Data Frameworks: Introduction to Scalable Applications & ML Challenges at Scale, Algorithms for Large-Scale Learning, Overview of Hadoop and Current Big Data Systems, Programming for Data Flow – Concepts & Differences, Apache Spark Basics – Vectors, Matrices, Spark ML Overview, Beyond Parallelization – Practical Big Data Applications

UNIT-II

Fast Data Applications and Messaging Systems: Anatomy of Fast Data Applications, SMACK Stack – Functional Decomposition, Message Backbone – Messaging Requirements, Data Ingestion, Low Latency & Fast Data, Message Delivery Semantics & Distribution of Messages.

UNIT-III

Compute Engines and Deployment for Fast Data: Compute Engines – Micro-Batch Processing, One-at-a-Time Processing, Engine Selection. Storage as Fast Data Border & Message Backbone as Transition Point. Sharing Stateful Streaming State. Data-Driven Microservices – State & Microservices. Deployment Environments – Containerization, Resource Scheduling, Apache Mesos, Kubernetes, Cloud Deployments.

REFERENCES:

1. Jan Kunigk, Ial Buss, Paul Wilkinson & Lars George," Architecting Modem Data Platforms", O"reilly, 2019.
2. Gerard Maas, Stavros Kontopoulos, Sean Glover, "Designing Fast Data Application Architectures", O'Reilly Media, Inc., June 2018.
3. Bill Chambers, Matei Zaharia "Spark- The definitive Guide", O'Reilly Media, Inc., June 2019.

M.Sc. (DATA SCIENCE) IV-SEMESTER SYLLABUS
MDS-405 : PAPER- V: PRACTICAL PAPER: CRYPTOGRAPHY & DATA MINING

Cryptography Practicals:

1. **Classical Encryption Techniques**
   a) Implement Caesar Cipher, Playfair Cipher, and Vigenère Cipher for encryption and decryption.
   b) Extend with a simple **transposition cipher**.
2. **Block Cipher Operations**
   a) Implement DES (or simplified DES) encryption and decryption.
   b) Demonstrate ECB, CBC, CFB, OFB, and Counter (CTR) modes with sample plaintext.
3. **Public-Key Cryptography (RSA)**
   a) Implement RSA key generation, encryption, and decryption.
   b) Verify with a short text message.
4. **Hashing and Authentication**
   a) Implement SHA-256 and MD5 hashing.
   b) Demonstrate HMAC for message authentication.

Data Mining Practical:

5. **Frequent Pattern Mining**
   a) Implement the Apriori algorithm for frequent itemset generation.
   b) Generate and display association rules with confidence and support.
6. **Classification**
   a) Implement Decision Tree Classifier using Gini Index or Entropy.
   b) Train and test it on a dataset (e.g., Iris or UCI dataset).
7. **Clustering**
   a) Implement K-Means clustering and visualize clusters.
   b) Compare with Hierarchical clustering (single-link / complete-link).
8. **Data Similarity and Visualization**
   a) Write a program to compute Euclidean, Cosine, and Jaccard similarity between data objects.
   b) Visualize data distribution using histograms, scatter plots, and boxplots.

M.Sc. (DATA SCIENCE) IV-SEMESTER SYLLABUS
MDS-406A : PAPER- VI : PRACTICAL PAPER E-I &E-II

ELECTIVE-I (A): SENTIMENTAL ANALYSIS

1. Implement a sentiment classification model to classify IMDB movie review dataset as positive or negative using machine learning techniques like Naïve Bayes and Logistic Regression.
2. Implement Python program to classify sentences as subjective/objective and then sentiment (positive/negative) on Movie review dataset split into sentences.
3. Implement Python program to build a sentiment lexicon using Dictionary-based and Corpus-based approaches to compare effectiveness on test sentences.
4. Implement Python program to Identify and analyze comparative opinions using Laptop/Mobile product reviews dataset.
5. Apply classification techniques to Detect fake/deceptive/spam reviews using Yelp Review Spam Dataset. (Hint: classification: Logistic Regression, Random Forest).

ELECTIVE-I (B): COMPUTER VISION

1. Write a Python program to simulate geometric transformations (translation, rotation, scaling) on an image. Demonstrate how a digital camera captures an image using pixel grids.
2. Implement point operations (brightness, contrast adjustment, negative image) and apply linear filtering (smoothing, sharpening). Perform Fourier Transform on an image and visualize its frequency spectrum.
3. Develop a Python program to detect edges (Canny/Sobel), corners (Harris/ORB), and match keypoints between two images. Demonstrate with object matching between two scenes.
4. Implement image segmentation techniques such as K-means clustering and active contours to separate objects from the background. Compare results with graph-based segmentation (Graph Cuts).
5. Build a simple object/face recognition system using feature extraction (HOG/SIFT/ORB) and classification (SVM/KNN). Evaluate recognition accuracy on a small dataset (e.g., ORL Face Dataset or CIFAR-10 subset).

ELECTIVE-I (C): MACHINE LEARNING OPERATIONS

1. Write a Python program to set up an MLOps environment by installing and configuring required tools such as Python, Git, Docker, and Jupyter Notebook. Demonstrate the creation of a virtual environment and initialization of a Git repository.
2. Implement DevOps principles for a machine learning workflow by developing a Python program for data preprocessing and model training. Track code using Git and demonstrate automation using a CI/CD pipeline.
3. Develop an MLOps workflow for data versioning and collaboration using tools such as DVC/Git LFS and GitFlow. Demonstrate dataset version control, data lineage tracking, and collaborative development.

4. Write a program to containerize and deploy a machine learning model using Docker. Demonstrate deployment on a cloud platform or container orchestration tool and manage multiple model instances.
5. Implement continuous integration and monitoring for an ML model by writing automated tests, tracking model performance metrics, detecting model drift, and deploying updated model versions through a CI/CD pipeline.

ELECTIVE-II (A): DATA STREAM MINING

1. Implement stream data mining using MOA (Massive Online Analysis) to analyze continuous data streams. Demonstrate stream mining concepts, assumptions, data characteristics, and evaluate the stream using standard testing frameworks.
2. Develop a stream mining application to handle concept drift by implementing change detection strategies such as adaptive sliding windows. Demonstrate how evolving data streams are detected and managed in real-time.
3. Implement Hoeffding Tree induction for data streams using the Hoeffding bound. Demonstrate tree construction, prediction strategies (Majority Class, Naïve Bayes leaves), and compare batch learning with data stream learning.
4. Design and evaluate Hoeffding Tree ensemble methods for evolving data streams. Demonstrate ensemble learning with realistic ensemble sizes and analyze performance under changing stream conditions.
5. Implement adaptive stream mining techniques using Adaptive Hoeffding Trees and ensemble methods such as Bagging with ADWIN. Evaluate model performance and adaptability on evolving data streams.

ELECTIVE-II (B): WEB MINING

1. Implement Association Rule Mining using the Apriori algorithm to generate frequent itemsets and association rules. Demonstrate rule generation with minimum support, confidence, and multiple minimum support thresholds.
2. Develop a program for Sequential Pattern Mining using algorithms such as GSP or PrefixSpan. Demonstrate pattern discovery and rule generation from sequential data.
3. Implement supervised and unsupervised learning techniques for web data mining by applying Decision Trees, Naïve Bayes for text classification, and K-means or Hierarchical Clustering. Analyze strengths and weaknesses of each method.
4. Design an information retrieval system using Boolean Model and Vector Space Model. Perform text preprocessing (stopword removal, stemming, inverted index) and evaluate retrieval performance using standard measures.
5. Implement web mining techniques for link analysis and web crawling by demonstrating PageRank or HITS algorithms, crawler strategies (BFS / focused crawling), and sentiment classification of web documents.

ELECTIVE-II (C): SCALABLE ARCHITECTURE

1. Develop a scalable data processing application to demonstrate challenges of machine learning at scale. Use a Big Data framework such as Hadoop or Apache Spark and explain data flow programming concepts.
2. Implement a large-scale learning workflow using Apache Spark ML by performing vector and matrix operations. Demonstrate how Spark enables scalable machine learning beyond simple parallelization.
3. Design a fast data application using a messaging system based on the SMACK stack architecture. Demonstrate data ingestion, message delivery semantics, and low-latency data processing.
4. Develop a streaming data processing application using micro-batch or one-at-a-time processing engines. Compare engine selection strategies and demonstrate state sharing in streaming applications.
5. Deploy a fast data application using containerization by demonstrating microservices architecture. Use Kubernetes or a cloud platform to manage resource scheduling and scalable deployment.

M.SC. (DATA SCIENCE) IV-SEMESTER

MDS-407: PAPER- VII: CAPSTONE PROJECT-II

Objectives:

1. To enhance practical and professional skills.
2. To familiarize tools and techniques of systematic Literature survey and documentation
3. To expose the students to industry practices and team work.
4. To encourage students to work with innovative and entrepreneurial ideas

Outcomes: Student will be able to:

1. Demonstrate the ability to synthesize and apply the knowledge and skills acquired in the academic program to real-world problems
2. Evaluate different solutions based on economic and technical feasibility
3. Effectively plan a project and confidently perform all aspects of project management
4. Demonstrate effective written and oral communication skills

Guidelines for Project:

1. Each Student has to do one standalone Statistical data analysis project in IV semester (Major project) (it should be different with III sem project) to familiarize the practical usage of all statistical techniques covered in UG & PG level using Statistical software's Python and R, from any one industry/ institution under a recognized Supervisor (supervisor must be eligible for teaching at PG level Statistics in any University as per UGC norms).
2. Suggested to collect Live large scale data set (may be primary or secondary source data set with minimum sample size 2000 with minimum 10 data variables with different measurement of scales of variables)
3. Each student has to submit the project reports in two copies (minimum 100 pages) hard bound copy along with the assignments follow the Ph.D. thesis norms as per Telangana University duly signed by the Students on Declaration, Certificate from Industry /institution and certified by recognized Supervisor supported by Plagiarism report (not exceeding 10% of similarity index).
4. The Project Report should fulfill the norms of statistical data analysis report. (It should contains (a) Literature collected related to the study and its review (research articles minimum 10) & Data Domain Description (b) National & International significance on the problem (c) Detailed data variable description (d) Data set objectives/ objectives of the study (e) Formation of statistical hypothesis (f) Data visualization techniques applied on each/combination of variables (g) Descriptive statistics (h) Exploratory Data Analysis on each variable understudy, (i) Advanced data analysis tools applied (minimum 5), (j) Results analysis and its interpretation. (k) sample data set (min 20-50), Python / R-programs written for implementation should be placed in appendix (l) Bibliography; Analysis report. Usage of Python programs should be presented in Appendix.
5. Projects will be evaluated by two subject experts by conducting Project viva-voce along with other practicals. May require presentation on the Project.
6. Project Marks will be awarded based on the Criteria in the viva-voce performance in (i) Project topic chosen and its significance (ii) Depth of the theoretical & technical soundness and practical Implementation, domain knowledge, proper organization of chapters (iii) seminar presentation & communication skills, role/ contribution of the student in the project etc.

Project Report guidelines:

1. Title Page in prescribed format
2. Declaration by the student in prescribed format
3. Certificate from the Industry / supervisor in prescribed format
4. Acknowledgments to supervisor and who helped in the project.
5. Contents in prescribed format

Chapters:

1. **Introduction** (Introduction, Motivation to topic, Significance and need of the study, Problem Statement, Objectives of the study, Chapter wise summary).
2. **Review of the Literature** (Introduction with list of authors contributed (collected from journals) related to topic, Each author method description (5-10 methods) and its comparison study
3. **Data Domain** (About Data set and its domain knowledge with Sample data description,)
4. **Statistical tools and software's used** (Basic and advanced tools used and minimum 5 advanced methods used and their procedures, Software's used, etc)
5. **Data visualizations (**Tools applicable, justification, program Outputs/ results analysis and interpretations and conclusions**)**
6. **Exploratory Data Analysis** Tools applicable, justification, program Outputs/ results analysis and interpretations and conclusions
7. **Model Building /Advanced data analysis techniques** Tools applicable, justification, program Outputs/ results analysis and interpretations and conclusions
8. **Conclusions & Future Scope**
9. **Appendix** (d. Program Code for Implementation, sample data set if any)
10. **Bibliography** (Minimum 30 articles)
11. **Plagiarism Report**