

M.Sc. Data Science
Course under Choice Based Credit System
M.SC. (Data Science) I- YEAR
I-SEMESTER

Paper Code	Title of the paper	No of Hours	Max. Marks	Credits
MDS-101	Mathematical Foundations for Data Science	4	30+70 =100	4
MDS-102	Design and Analysis of Algorithms	4	30+70 =100	4
MDS-103	Software Engineering	4	30+70 =100	4
MDS-104	A: Principles of Data Science B: Java Programming	4	30+70 =100	4
MDS-105	Design and Analysis of Algorithms Lab	4	50	2
MDS-106	A: Principle of Data Science Lab B: Java Programming Lab	4	50	2
	Total	24	500	20

II-SEMESTER

Paper Code	Title of the paper	No of Hours	Max. Marks	Credits
MDS-201	Statistical Inference	4	30+70 =100	4
MDS-202	Data Visualization Techniques	4	30+70 =100	4
MDS-203	Artificial Intelligence	4	30+70 =100	4
MDS-204	Advanced Machine Learning Techniques	4	30+70 =100	4
MDS-205	Statistical Inference & Data Visualization Lab	4	50	2
MDS-206	Advanced Machine Learning Lab	4	50	2
	Total	24	500	20

Signature
 Head
 of Computer Science
 College for Women
 Hyderabad - 500095

Signature
 Prof. N. Ch. Bhattacharyulu
 M.Sc., Phil., Ph.D.
 Department of Statistics
 University College of Science
 Osmania University, Hyd-07

Signature
 HEAD
 Statistics Dept.
 HYDERABAD-7

M.Sc. (Data Science) -I Year -I Semester

Theory: 4Hours/Week

Internal Marks: 30

External Marks: 70

Credits: 4

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

UNIT – I

Vector Spaces: 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, rank- nullity theorem, change of basis and similar matrices Inner-product spaces, orthogonal sets and bases, Orthogonal Projection, Gram-Schmidt orthogonalization process

UNIT-II

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,

UNIT – III

Eigenvalue Problems: Characteristic roots and vectors, Caley-Hamilton theorem, Diagonalization of a Matrix, algebraic and geometric multiplicity of a characteristic root and spectral decomposition of a real symmetric matrix., Singular value Decomposition. Gauss-Jordan-LU decomposition, Singular Value Decomposition,

UNIT – IV

Quadratic forms: Real quadratic forms (QFs), reduction and classification of QFs, index and signature. Simultaneous reduction of two QFs. Extreme form of a QF. Moore-Penrose and generalized inverses and their properties.

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. Graybill, F.A. (1983) : Matrices with applications in statistics, 2nd ed, Wadsworth.
4. Rao, C.R., Mithra, S.K. (1971): Generalised inverse of matrices and its applications, John Wiley & Sons Inc.
5. Rao, C.R. and Bhimasankaram, P. (1992): Linear algebra. TMH.

M.Sc. (Data Science) -I Year -I Semester

Theory: 4Hours/Week

Internal Marks: 30

External Marks: 70

Credits: 4

MDS-102: 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; QuickSort, Selection sort, Strassen's Matrix Multiplication, Convex Hull.

UNIT-II

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

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

UNIT-III

Basic Traversal and Search Techniques: Techniques for Binary Trees, Techniques for Graphs, Connected Components and Spanning Trees, Biconnected Components and DFS.

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

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

UNIT -IV

NP-Hard and NP-Complete Problems: Basic Concepts, Cook's Theorem, NP-Hard. 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, CE Leiserson, R.L Rivert, C Stein, (2010): Introduction to Algorithms, 3/e, PHI.
3. R. Pannarselvam (2007): Design and Analysis of Algorithms, PHI.
4. Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

M.Sc. (Data Science) -I Year -I Semester

Theory: 4Hours/Week Internal Marks: 30 External Marks: 70 Credits: 4

MDS-103: PAPER- III: SOFTWARE ENGINEERING

UNIT – I

Software and Software Engineering: The Nature of Software, The Unique Nature of WebApps, Software Engineering, Software Process, Software Engineering Practice, Software Myths.

Process Models: A Generic Process Model, Process Assessment and Improvement, Prescriptive Process Models, Specialized Process Models, The Unified Process, Personal and Team Process Models, Process Terminology, Product and Process.

UNIT – II

Requirements Analysis And Specification: Requirements Gathering and Analysis, Software Requirement Specification (SRS), Formal System Specification.

Software Design: Overview of the Design Process, Characterization of Design, Cohesion and Coupling, Layered Arrangement of Modules, Approaches to Software Design.

Function-Oriented Software Design: Overview of SA/SD Methodology, Structured Analysis, Developing the DFD Model of a System, Structured Design, Detailed Design, Design Review, over view of Object Oriented design.

User Interface Design: Characteristics of Good User Interface, Basic Concepts, Types of User Interfaces, Fundamentals of Component-based GUI Development, A User Interface Design Methodology.

UNIT – III

Coding And Testing: Coding, Code Review, Software Documentation, Testing, Unit Testing, Black-Box Testing, White-Box Testing, Debugging, Program Analysis Tool, Integration Testing, Testing Object-Oriented Programs, System Testing, Some General Issues Associated with Testing.

Software Reliability And Quality Management: Software Reliability, Statistical Testing, Software Quality, Software Quality Management System, ISO 9000, SEI Capability Maturity Model.

UNIT – IV

Software Maintenance: Maintenance, Re-Engineering, Forward engineering and Reverse Engineering.

Computer Aided Software Engineering (CASE): CASE and its Scope, CASE Environment, CASE Support in Software Life Cycle, Characteristics of CASE Tools, Architecture of a CASE Environment, Second Generation CASE Tools,

REFERENCE BOOKS

1. Software Engineering A practitioner's Approach, Roger S. Pressman, Seventh Edition McGraw Hill International Edition.
2. Fundamentals of Software Engineering, Rajib Mall, Third Edition, PHI.
3. Software Engineering, Ian Sommerville, Ninth edition, Pearson education
4. Software Engineering : A Primer, Waman S Jawadekar, Tata McGraw-Hill, 2008
5. Software Engineering, A Precise Approach, Pankaj Jalote, Wiley India, 2010.
6. Software Engineering, Principles and Practices, Deepak Jain, Oxford University Press.
7. Software Engineering: Abstraction and modeling, Diner Bjorner, Springer International edition, 2006

Snub

Prof. N. Ch. Bhattacharyulu
M.Sc., M.Phil., Ph.D.
Department of Statistics
University College of Science

HEAD
Statistics Dept.
O.U. HYDERABAD

M.Sc. (Data Science) -I Year -I Semester

Theory: 4Hours/Week Internal Marks: 30 External Marks: 70 Credits: 4

MDS-104(A): PAPER- IV(A): PRINCIPLES OF DATA SCIENCE

(Mandatory for B.Sc. Computer Science Stream)

UNIT – I

Introduction to Data: Data types, Measurement of scales, understanding data with descriptive statistics and with data visualization and data pre-processing steps. Data transformations (Standardize, Normalize, converting data from one scale to other scales). Parametric & Non-Parametric tests (z -, χ^2 , t -, F -tests; Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test) and Feature selection methods.

UNIT – II

Introduction to Data Modelling: Review of the modelling process, Concepts of Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and Leave out one approaches), Supervised and Un-supervised Modelling, Model Performance for classification techniques for qualitative and Quantitative data, Model improvement and saving models for future use.

UNIT-III

Supervised learning algorithms: Classification algorithms, Multiple Linear regression, Multinomial logistic, Support vector Machine, Random-forest, Ada Boost, Ensemble methods. Decision Tree, Perceptron learning, Naïve Bayes classifier.

UNIT-IV

Unsupervised learning Algorithms: Introduction to Cluster analysis, Clustering techniques, Agglomerative Hierarchical cluster techniques, K-means, K-medoid, KNN.

REFERENCE BOOKS

1. Foster Provost and Tom Fawcett (2021): Data Science for Business, O'REILLY Publications
2. Henrik Brink, Joseph W. Richards. Mark Fetherolf (2012): Real World Machine Learning, Manning Publications.

M.Sc. (Data Science) -I Year -I Semester

Theory: 4Hours/Week

Internal Marks: 30

External Marks: 70

Credits: 4

MDS-104(B): PAPER- IV(B): JAVA PROGRAMMING

(Mandatory for B.Sc. Data Science stream)

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.

UNIT – II

Overview of Networking: Working with URL, Connecting to a Server, Implementing Servers, Serving multiple Clients, Sending EMail, 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 – III

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 – IV

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. Herbert Schildt, Java Complete Reference
3. Cay S. Horstmanns, Gray Coronell, Core Java Vol. II – Advanced Features
4. Sharanam Shah, Vaishali Shah, Java EE 7 for Beginners

Snub
Dept. of Computer Science
University College for Women
Hyderabad - 500095

Prof. N. Ch. Bhattacharyulu
M.Sc., M.Phil., Ph.D.
Department of Statistics
University College of Science
Osmania University

9.75
HEAD
Statistics Dept.
OU, Hyderabad
Page

Women's University, Kott, Hyderabad
M.Sc. (Data Science) -I Year -I Semester

Practical: 4Hours/Week Internal Marks30: External Marks: 20 Credits: 2

MDS-105 (LAB-1): PAPER- V: DESIGN & ANALYSIS OF ALGORITHMS LAB

List of Experiments (Using Python)

1. Write a program for sorting the given list using Insertion Sort.
2. Write a program for sorting the given list using Selection Sort.
3. Write a program for sorting the given list using MergeSort.
4. Write a program for sorting the given list using Quick Sort.
5. Write a program for sorting the given list using Heap Sort.
6. Write a program to find the given number in a list using Sequential Search.
7. Write a program to find the given number in a list using BinarySearch.
8. Write a program to find the minimal spanning tree for a weighted graph using Kruskal's and prims Algorithms.
9. Write a program to find the shortest path in a weighted graph using Dijkstra's Algorithm.
10. Write a program to solve using dynamic programming technique for Travelling sales man problem.
11. Write a program to solve using dynamic programming technique for Multistage Graph problem using All-Pairs Shortest Paths (Warshal),
12. Write a program to solve using dynamic programming technique for Multistage Graph problem using Single-Source Shortest Paths (Bellman ford),
13. Write a program to solve using dynamic programming technique for Multistage Graph problem using Optimal Binary Search Trees
14. Write a program to solve using dynamic programming technique for Travelling sales man problem using Optimal Binary Search Trees.
15. Write a program to solve Knapsack problem using Back tracking

Smruti
Head
of Computer Science
City College for Women
Hyderabad - 500099

Prof. N. Ch. Bhatracharyulu
M.Sc., M.Phil., Ph.D.
Department of Statistics
University of Science

AD
Statistics Dept.
O.M. Hyderabad-7

M.Sc. (Data Science) -I Year -I Semester

Practical: 4Hours/Week Internal Marks:30 External Marks: 20 Credits: 2

MDS-106(A) (LAB-2): PAPER- VI: PRINCIPLES OF DATA SCIENCE LAB

List of Experiments (Using Python)

1. Data Preprocessing
2. Write a Python code for the implementation of Classification techniques for the data sets:
 - a) Multiple Linear regression,
 - b) Multinomial logistic,
 - c) Support vector Machine,
 - d) Random-forest,
 - e) Ada Boost/ XG Boost.
 - f) Decision Tree,
 - g) Naïve Bayes classifier.
3. Write a Python code for the implementation of Cluster analysis for the data sets using:
 - a) Agglomerative Hierarchical cluster techniques,
 - b) K-means,
 - c) KNN.

Head
Dept. of Computer Science
University College for Women
Hyderabad - 500095

Prof. N. Ch. Bhadracharyulu
M.Sc., M.Phil., Ph.D.
Department of Statistics
University College of Science
Hyderabad - 500095

Head
Statistics Dept.
U.C.W. Hyderabad-7

Women's University, Kott, Hyderabad
M.Sc. (Data Science) -I Year -I Semester

Practical: 4Hours/Week Internal Marks:30 External Marks: 20 Credits: 2

MDS-106(B) (LAB-2): PAPER- VI: JAVA PROGRAMMING LAB

List of Experiments

1. Write a program to demonstrate Single Inheritance
2. Write a Program to demonstrate Multiple Inheritance (Interfaces & Abstract Class).
3. Write a program for working with User Defined Packages
4. Write a program for working with String & String Buffer Class
5. Write a program to create a registration form with different controls, menus and demonstrate event handling.
6. Write a program to copy data from one file to another file.
7. Write a program to merge contents of two files and display output on console Create GUI to demonstrate swing components using student registration form.
8. Write a program for working Exceptions (User-Defined & Inbuilt)
9. Write a program to work with Thread class and Runnable Interface.
10. Write a program for demonstrating Serialization
11. Write a program for creating Table for inserting, selecting, and deleting the records.
12. Develop Servlet application to accept HTNO of a student from client and display the memorandum of marks from the server
13. JSP programs a. Create a JSP page that prints temperature conversion (from Celsius to Fahrenheit) chart b. Create a JSP page to print current date and time c. Create a JSP page to print number of times page is referred after the page is loaded.
14. Write a simple JSP application to demonstrate the use of implicit object (at least 5).
15. Develop a JSP Application to accept Registration Details from the user and store database table.

WOMEN'S UNIVERSITY, KOU, HYDERABAD
M.Sc. (Data Science) -I Year -II Semester

Theory: 4Hours/Week Internal Marks: 30 External Marks: 70 Credits: 4

MDS-201: PAPER- I: STATISTICAL INFERENCE

UNIT-I

Estimation Theory: Desirable properties of a good estimator: Unbiasedness, consistency, efficiency and sufficiency - 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, Kernel and examples. Interval estimation, confidence level CI using pivots and shortest length CI and example problems.

UNIT-III

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-IV

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. Prior and Posterior 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)

Scanned by
Head
of Computer Science
City College for Women
Hyderabad - 500095

Prof. N. Ch. Bhattacharyulu
M.A. Phil., Ph.D.
Department of Statistics
University of Science

Signature
NEAD
Statistical Dept.
WU, HYDERABAD

YOUNG'S UNIVERSITY, KOLH, HYDERABAD
M.Sc. (Data Science) -I Year -II Semester

Theory: 4Hours/Week Internal Marks: 30 External Marks: 70 Credits: 4

MDS-202: PAPER- II: DATA VISUALIZATION TECHNIQUES

UNIT-I

Introduction to Data Visualization: Introduction, importance of Data visualization, Human perception – Methodology – Seven Stages of Data Visualization - Data Visualization Tools. Visualizing Data: Mapping Data onto Aesthetics – Visualizing Amounts - Visualizing Distributions: Histograms and Density Plots – Visualizing Propositions: – Visualizing Associations: Among Two or More Quantitative Variables – Visualizing Time Series and Other Functions of an Independent Variable – Trends – Visualizing Geospatial Data.

UNIT-II

Basic Applied Visualizations: Data Visualization: One dimensional (Pictogram, Pie Chart, Bar Chart.), two-dimensional (Histogram, Line plot, frequency curves& polygons, ogive curves, Scatter Plot,) and other data visualization techniques. Gantt Chart, Heat Map, Box and Whisker Plot, Waterfall Chart, Area Chart, Stacked Bar Charts - Sub Plots – Matplotlib, Seaborn Styles, Box plot - Density Plot - - Tree map - Graph Networks.

UNIT-III

Interactive Visualisation: Introduction to D3 - Fundamental Technology: The Web – HTML – DOM – CSS – JavaScript – SVG. D3 Setup – Generating Page Elements – Binding Data - Drawing with data – Scales: Domains and Ranges – Normalization – Creating a Scale – Scaling the Scatter Plot – Other Methods and Other Scales. Axes – Modernizing the Chart – Update the Data – Transition – Updates – Interactivity. Timeline, Highlight Table, Bullet Graph, Choropleth Map, Word Cloud, Path diagram, Network Diagram, Correlation Matrices.

UNIT-IV

Principles of Information Visualization: Visual Perception and Cognition, Gestalt's Principles, Tufte's Principles, Applications of Principles of Information Visualization, Dashboard Design.

REFERENCE BOOKS

1. Julie Steele, Noah Iliinsky (2010); Beautiful Visualization Publisher(s): O'Reilly Media
2. Edward R. Tufte (2016): "The Visual Display of Quantitative Information" by
3. Kieran Healy(2016): Data Visualization: A Practical Introduction,
4. Claus O. Wilke (2020) Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures., O'reilly
5. Stephanie D.H. Evergreen(2020): Effective Data Visualization: The Right Chart for the Right Data

Signature
Computer Science
College for Women
Hyderabad - 500095

Prof. N. Ch. Bhattacharyulu
M.Sc., M.Phil., Ph.D.
Department of Statistics
College of Science

Signature
HEAD
Statistics Dept.
O.M. HYDERABAD-7

WOMEN'S UNIVERSITY, KOU, HYDERABAD
M.Sc. (Data Science) -I Year -II Semester

Theory: 4Hours/Week Internal Marks: 30 External Marks: 70 Credits: 4

**MDS-203: PAPER- III: ARTIFICIAL INTELLIGENCE
UNIT - I**

Introduction: History Intelligent Systems, Foundations of Artificial Intelligence, Sub areas of AI, 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.

UNIT - IV

Advanced Knowledge Representation Techniques: Case Grammars, Semantic Web. Natural Language Processing: Introduction, Sentence Analysis Phases, Grammars and Parsers, Types of Parsers, Semantic Analysis, Universal Networking Knowledge.

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.

Computer Science
Head
for Women
500095

Prof. N. Ch. Bhattacharyulu
M.Phil, Ph.D.
Statistics

HEAD
Statistics Dept.
HYDERABAD-7

M.Sc. (Data Science) -I Year -II Semester

Theory: 4Hours/Week Internal Marks: 30 External Marks: 70 Credits: 4
MDS-204: PAPER-IV: ADVANCED MACHINE LEARNING TECHNIQUES

UNIT – I

Multivariate Data Analysis Techniques: Path analysis, Correspondence analysis, Multidimensional Scaling, Feature extraction and Feature selection techniques, Inter and intra class distance measures, Probabilistic distance measures, Principal component analysis, Factor analysis, Canonical Correlations and Canonical Variables. Conjoint Analysis.

UNIT - II

Classification Techniques-I: Linear classifies, Multiple Linear regression, Logistic regression, Linear Discriminant Function (for binary outputs) with minimum squared error, Linear discriminant function using likelihood ratios based on Multivariate normal populations, Bayes Mis-classification; Naïve Bayes classifier, Support Vector Machines,

UNIT – III

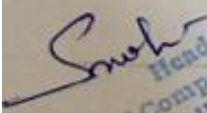
Classification Techniques-II: Decision Tree algorithms, Random Forest algorithm, Bagging, Gradient boosting, Ada-Boosting and XG-Boosting algorithm, KNN algorithm, Market-Basket Analysis.

UNIT – IV

Cluster Analysis: Introduction, similarities and dissimilarities, Hierarchical clustering, Single linkage method, k-means and k-Nearest Neighbourhood (KNN) clustering, Categorical Data Analysis techniques and their applications; Time series forecasting: ARIMA model, ARCH & GARCH model.

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.


Head
of Computer Science
College for Women
Hyderabad - 500095

Prof. N. Ch. Bhattacharyulu
M.Sc., Ph.D.
Department of Statistics
University College of Science
Osmania University Hyd-07


HEAD
Statistics Dept.
HYDERABAD

M.Sc. (Data Science) -I Year -II Semester

Practical: 4Hours/Week

Internal Marks: 30

External Marks:20

Credits: 2

MDS-205(LAB-1): PAPER- V: DATA VISUALISATION & INFERENCE LAB

List of Experiments (using Python)

1. Drawing One dimensional diagrams (Pictogram, Pie Chart, Bar Chart,).
2. Drawing two-dimensional (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot)
3. Drawing 3D and other data visualization techniques.
4. Drawing Gantt Chart, Heat Map, Box - Whisker Plot, Waterfall Chart, Area Chart, Stacked Bar Charts –
5. Drawing Density Plot, Bullet Graph, Choropleth Map, Tree map, Path diagram, Network Diagram, Correlation Matrices.
6. Generation of Jackknife and Bootstrap samples and its parameter estimation.
7. Computation of Confidence Interval estimation of parameters.
8. Small and Large sample tests (for Mean(s), Standard deviation/variance(s), Proportion(s)).
9. Non parametric tests: One and two sample tests (Kolmogorov Smirnov, Kruskal Wallis & Friedman test, Kendal's tau, Ansari broadly tests.)
10. Generation of random numbers for Uniform, Normal, Exponential, Cauchy and Poisson Distributions.
11. Bayesian estimation of parameters, (using Metropolis Hasting and Gibbs Sampler).

Snob
Head
of Computer Science
College for Women
Hyderabad - 500036

Prof. N. Ch. Bhattacharyulu
Prof. N. Ch. Bhattacharyulu
M.Phil., Ph.D.
Department of Statistics
University College of Science

eg: 18
HEAD
Statistics Dept.
O.M. HYDERABAD-9

With effect from the academic year 2023-2024
Department of Computer Science
TELANGANA MAHILA VISHWAVIDYALAYAM
Women's University, Koti, Hyderabad
M.Sc. (Data Science) -I Year -II Semester

Practical: 4Hours/Week Internal Marks:30 External Marks: 20 Credits: 2

MDS-206: (LAB-2): PAPER- VI: ADVANCED MACHINE LEARNING LAB

List of Experiments

1. Implement and demonstrate the use of set of training data samples. Read the training data from a .CSV file.
2. Write a program to demonstrate the working of the decision tree-based ID-3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
3. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
4. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
5. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.
6. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.
7. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
8. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
9. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Signature
Head
Department of Computer Science
T.M.V. College for Women
Hyderabad - 500005

Signature
Prof. N. Ch. Bhattacharyulu
M.Sc., M.Phil., Ph.D.
Department of Statistics
University College of Science

Signature
HEAD
Statistics Dept
T.M.V. HYDERABAD-5

M.Sc.Data Science
II Year Course Curriculum
(w.e.f. Academic Year 2023-24)
III-SEMESTER

Paper Code	Title of the paper	Credits	# Hours	Max. Marks
MDS-301	Deep Learning Techniques	4	4	100
MDS-302	Machine Learning Operations	4	4	100
MDS-303 (Elective-I)	a) Data Mining	4	4	100
	b) Text Data Analytics			
	c) Enterprise Architecture			
MDS-304 (Elective-II)	a) Optimization techniques	4	4	100
	b) Business Intelligence			
	c) Computer Vision			
MDS-305	Deep Learning Techniques Lab	2	4	50
MDS-306	Machine Learning Operations Lab	2	4	50
No of Credits		20	24	500

IV-SEMESTER

Paper Code	Title of the paper	Credits	# Hours	Max. Marks
MDS-401	Research Methodology	4	4	100
MDS-402	Cloud Computing	4	4	100
MDS-403 (Elective-I)	a) Data Stream Mining	4	4	100
	b) Sentimental Analysis			
	c) Scalable Architecture			
MDS-404	Project	8	16	200
No of Credits		20	28	500

[Signature]
 Chair-Person
 B.O.S. in Statistics
 U.C.S.
 Osmania University
 ABAD-7

[Signature]
 Professor, Dept. of Statistics
 Professor, Dept. of SCIENCE
 UNIVERSITY COLLEGE OF
 UNIVERSITY

[Signature]
 HEAD
 Department of Computer Science
 Telangana Mahila Viswavidyalaya
 (Women's University)

MDS-301

M.Sc. Data Science III-Semester
DEEP LEARNING TECHNIQUES

Credits:4

Instruction	4Lhrs per week	Duration of SEE	3hours
CIE	30marks	SEE	70marks

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.

UNIT – III

Unsupervised learning Algorithms: Hebbian Learning, Competitive learning, Self-Organizing Maps, SOM algorithm, properties of feature map, computer simulations, Vector quantization, Learning vector quantization. Radial Basis Function Networks, Approximation properties of Radial Basis Function Networks. Boltzman Machine, Hopfield model.

UNIT – IV

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. Fausett, L. (1994). Fundamentals of Neural Networks. New York: Prentice Hall.
7. Box and Jenkins: Time Series analysis, Springer
8. Brockwell, P.J., and Davis, R.A.: Time Series : Theory and Methods (Second Edition), Springer-Verlag.

Smit
HEAD
Department of Computer Science
Telangana Mahila Viswavidyalaya
Women's University
Koti, Hyderabad-500 095, T.S.

Prof. Dr. S. S. S.
Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCE
Osmania University

MDS-302

M.Sc. Data Science III-Semester
MACHINE LEARNING OPERATIONS

Credits:4

Instruction
CIE

4Lhrs per week
30marks

Duration of SEE
SEE

3hours
70marks

UNIT – I

Introduction to ML Operations: Overview of ML Ops and its significance in the industry, Challenges in deploying and managing machine learning models, Principles of DevOps and its application to ML Ops, Introduction to cloud platforms and tools for ML Ops.

UNIT – II

Data Management and Version Control: Data versioning and management in ML projects, Implementing reproducibility and data lineage, Git and version control for ML models and pipelines, Collaborative development workflows for ML Ops.

UNIT – III

Model Deployment and Monitoring: Model packaging and containerization, Deploying models to production environments, Infrastructure orchestration and scaling for ML workloads, Monitoring model performance and managing drift

UNIT – IV

Continuous Integration and Delivery for ML: Automated testing and validation of ML models, Continuous integration pipelines for ML Ops, Continuous deployment and release strategies, Feedback loops and continuous improvement in ML Ops

REFERENCE

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


Chair-Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCE
Osmania University
HYDERABAD-07, TS.


Chair-Person
Department of Computer Science
Osmania Mahila Viswavidyalaya
Women's University
Koti, Hyderabad-500 082

MDS-302

M.Sc. Data Science III-Semester
MACHINE LEARNING OPERATIONS

Credits:4

Instruction
CIE

4Lhrs per week
30marks

Duration of SEE
SEE

3hours
70marks

UNIT – I

Introduction to ML Operations: Overview of ML Ops and its significance in the industry, Challenges in deploying and managing machine learning models, Principles of DevOps and its application to ML Ops, Introduction to cloud platforms and tools for ML Ops.

UNIT – II

Data Management and Version Control: Data versioning and management in ML projects, Implementing reproducibility and data lineage, Git and version control for ML models and pipelines, Collaborative development workflows for ML Ops.

UNIT – III

Model Deployment and Monitoring: Model packaging and containerization, Deploying models to production environments, Infrastructure orchestration and scaling for ML workloads, Monitoring model performance and managing drift

UNIT – IV

Continuous Integration and Delivery for ML: Automated testing and validation of ML models, Continuous integration pipelines for ML Ops, Continuous deployment and release strategies, Feedback loops and continuous improvement in ML Ops

REFERENCE

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


Chair-Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCE
Osmania University
HYDERABAD-07, TS.


Chair-Person
Department of Computer Science
Telangana Mahila Viswavidyalaya
(Women's University)
Koti, Hyderabad-500 025

Department of Computer Science
Telangana Mahila Viswavidyalaya, Hyderabad
M.Sc. Data Science III-Semester

MDS-304
(Elective-II)

(A): OPTIMIZATION TECHNIQUES

Credits:4

Instruction
CIE

4Lhrs per week
30marks

Duration of SEE
SEE

3hours
70marks

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 & twophase 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

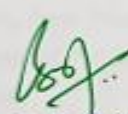
Integer Programming Problem: Gomory's cutting plane Algorithm for pure and mixed IPP Branch and bound Technique. Networks: Basic concepts of Networks constraints; Construction of Network and critical path; PERT and CPM; Network flow problems. Time Cost Analysis.

UNIT-IV

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 Lieberman,G.J.(1962): Introduction to Operations Research; Holdon Day


Chair-Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCE
Osmania University
HYDERABAD-07, T.S.


HEAD
Department of Computer Science
Osmania Mahila Vidyapeeth
(Women's University)
Koti, Hyderabad-500 090, T.S.

Department of Computer Science
Telangana Mahila Viswavidyalaya, Hyderabad
M.Sc. Data Science III-Semester
DEEP LEARNING TECHNIQUES LAB

MDS-305
(LAB-I)

Credits:2

Instruction 3P hrs per week
CCE(Internal) 30marks

Duration of SEE
External

3hours
20marks

LIST OF EXPERIMENTS (Using Python / R)

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


Chair-Person Professor, Dept. of Statistics
B.O.S. in Statistics
U.C.S. UNIVERSITY COLLEGE OF SCIENCE
Osmania University
HYDERABAD-500 007


HEAD
Department of Computer Science
Telangana Mahila Viswavidyalaya
(Women's University)
Koti, Hyderabad-500 095.

Department of Computer Science
Telangana Mahila Viswavidyalaya, Hyderabad
M.Sc. Data Science III-Semester
MACHINE LEARNING OPERATIONS LAB

MDS-306
(LAB-II)

Credits:2

Instruction 3P hrs per week
CCE(Internal) 30marks

Duration of SEE
External

3hours
20marks

1: Setting up a ML Ops Environment:

- * Install necessary tools (e.g., Python, Jupyter Notebook, Git, Docker).
- * Configure a cloud platform (e.g., AWS, GCP, Azure) for ML Ops.
- * Create a virtual environment for project isolation.

2: Implementing DevOps Principles:

- * Write a simple Python script to preprocess and train a machine learning model.
- * Create a version control repository (e.g., GitHub) to track changes.
- * Automate the build and deployment process using a CI/CD pipeline (e.g., Jenkins, GitLab CI).

3: Data Versioning and Management:

- * Store data in a version control system (e.g., DVC, Git LFS).
- * Implement data lineage tracking to understand data transformations.
- * Experiment with different data preprocessing techniques and track their impact on model performance.

4: Collaborative Development Workflows:

- * Create a shared project repository and collaborate with team members.
- * Use code review and merge requests to ensure quality and consistency.
- * Implement branching strategies (e.g., GitFlow) for managing different development phases.

5: Model Packaging and Containerization:

- * Create a Docker image containing the trained model and dependencies.
- * Deploy the containerized model to a cloud platform (e.g., Kubernetes, AWS ECS).
- * Use a container orchestration tool to manage multiple model instances.

6: Monitoring Model Performance:

- * Set up metrics to track model performance (e.g., accuracy, precision, recall).
- * Implement anomaly detection to identify unexpected changes in model behavior.
- * Use A/B testing to compare different model versions.


Chair Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCE
Osmania University
HYDERABAD-07, TS.


Dean
Department of Computer
Telangana Mahila Viswavidyalaya
(Women's University)
Koti, Hyderabad-500004

7: Automated Testing and Validation:

- Write unit tests for model components (e.g., preprocessing, training, prediction).
- Implement integration tests to verify the entire ML pipeline.
- Use automated validation techniques (e.g., data quality checks, model drift detection).

8: Continuous Integration and Deployment:

- Integrate automated testing into the CI/CD pipeline.
- Configure continuous deployment to automatically deploy new model versions.
- Implement feedback loops to continuously improve model performance.

M.Sc. Data Science IV-Semester

MDS-401

RESEARCH METHODOLOGY

Credits:4

Instruction

4Lhrs per week

Duration of SEE

3hours

CIE

30marks

SEE

70marks

UNIT - I

Research Methodology: Objectives and Motivation of Research, Types of Research, research approaches, Significance of Research, Research Methods Verses Methodology, Research Process, Criteria of Good Research, Problems Encountered by Researchers in India, Benefits to the society in general. Defining the Research Problem: Selection of Research Problem, Necessity of Defining the Problem, Research ethics.

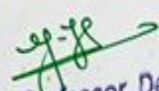
UNIT - II

Literature Survey: Importance and purpose of Literature Survey, Sources of Information, Assessment of Quality of Journals and Articles, Need of Review, Guidelines for Review, Record of Research Review.

UNIT - III

Research Design: Meaning of Research Design, Need of Research Design, Feature of a Good Design, Important Concepts Related to Research Design, Different Research Designs (exploratory, descriptive, experimental), Basic Principles of Experimental Design, Developing a Research Plan, Steps in sample design, types of sample designs.


Chair-Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCES
Osmania University
HYDERABAD-07, TS.


HEAD
Department of Computer Science
Telangana Mahila Viswavidyalaya
(Women's University)
Koti, Hyderabad-500 093, T

UNIT - IV

Report writing: Meaning of interpretation, layout of research report, Types of reports, Mechanism of writing a report. Research Proposal Preparation: Writing a Research Proposal and Research Report, Writing Research Grant Proposal, Paraphrasing & Plagiarism. Process of Writing a research paper; Indexing, Citation of sources; Writing first draft of thesis, Revising/ Editing - The final draft and proof reading; Research Paper Publication: Reputed Journals - National/International - ISSN No, No. of volumes, Scopus Index/UGC Journals - Free publications;

REFERENCE

1. C.R Kothari, Research Methodology, Methods & Technique; New Age International Publishers, 2004
2. R. Ganesan, Research Methodology for Engineers, MJP Publishers, 2011
3. Y.P. Agarwal, Statistical Methods: Concepts, Application and Computation, Sterling Publications Pvt. Ltd., New Delhi, 2004.



Chair-Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7



Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCES
Osmania University
HYDERABAD-07, TS.



HEAD
Department of Computer Science
Telangana Mahila Vidyapeeth
(Women's University)
Koti, Hyderabad-500 099

Department of Computer Science
Telangana Mahila Viswavidyalayam, Hyderabad
M.Sc. Data Science IV-Semester

MDS-402

CLOUD COMPUTING

Credits:4

Instruction
CIE

4Lhrs per week
30marks

Duration of SEE
SEE

3hours
70marks

UNIT-I

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

UNIT -II

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

UNIT-III

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- V

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

REFERENCE

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.


Chair Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor
Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCE
Osmania University
HYDERABAD-7. TS.


HEAD
Department of Computer Science
Telangana Mahila Viswavidyalayam
(Women's University)
Koti, Hyderabad-500 095, T.S.

Department of Computer Science
Telangana Mahila Vidyalyayam, Hyderabad
M.Sc. Data Science IV-Semester

MDS-403
(Elective-I)

(B): SENTIMENT ANALYSIS

Credits:4

Instruction
CIE

4Lhrs per week
30marks

Duration of SEE
SEE

3hours
70marks

UNIT-I

Introduction to Sentiment Analysis Introduction: Sentiment Analysis Applications - Sentiment Analysis Research - Sentiment Analysis as Mini NLP. The Problem of Sentiment Analysis: Definition of Opinion - Definition of Opinion Summary - Affect, Emotion, and Mood - Different Types of Opinions - Author and Reader Standpoint. Document Sentiment Classification: Supervised Sentiment Classification - Unsupervised Sentiment Classification - Sentiment Rating Prediction - Cross-Domain Sentiment Classification - Cross-Language Sentiment Classification - Emotion Classification of Documents.

UNIT-II

Subjectivity Classification and Challenges: Sentence Subjectivity and Sentiment Classification: Subjectivity - Sentence Subjectivity Classification - Sentence Sentiment Classification - Dealing with Conditional Sentences - Dealing with Sarcastic Sentences - Cross-Language Subjectivity and Sentiment Classification - Using Discourse Information for Sentiment Classification - Emotion Classification of Sentences.

UNIT-III

Sentiment Lexicon generation and Summarization: Sentiment Lexicon Generation: Dictionary-Based Approach - Corpus-Based Approach - Desirable and Undesirable Facts. Analysis of Comparative Opinions: Problem Definition - Identify Comparative Sentences - Identifying the Preferred Entity Set - Special Types of Comparison - Entity and Aspect Extraction. Opinion Summarization and Search: Aspect-Based Opinion Summarization - Enhancements to Aspect-Based Summary - Contrastive View Summarization - Traditional Summarization - Summarization of Comparative Opinions - Opinion Search - Existing Opinion Retrieval Techniques. Mining Intentions: Problem of Intention Mining - Intention Classification - Fine-Grained Mining of Intentions.


Department of Computer Science
Telangana Mahila Vidyalyayam
(Women's University)
Koti, Hyderabad-500 095, T.S.

UNIT-IV

Identifying intention, fake and quality of opinion: Detecting Fake or Deceptive Opinions:
Different Types of Spam - Supervised Fake Review. Detection - Supervised Yelp Data
Experiment - Automated Discovery of Abnormal Patterns - Model
Based Behavioral Analysis - Group Spam Detection - Identifying Reviewers with Multiple User
ids - Exploiting Business in Reviews - Some Future Research Directions. Quality of Reviews:
Quality Prediction as a Regression Problem - Other Methods - Some New Frontiers.

REFERENCE

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


Chair-Person
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCE
Osmania University
HYDERABAD-47, TS.


HEAD
Department of Computer Science
Telangana Mahila Viswavidyalayam
(Women's University)
Koti, Hyderabad-500 095, T.S.

MDS-404

**M.Sc. Data Science IV-Semester
PROJECT**

Credits: 8

Instruction 16L hrs per week
CIE(Internal) 100 marks

Duration of SEE
External

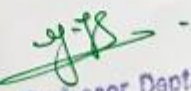
3hours
100 marks

1. Each student carries out a project individually.
2. The aim of project work is to develop solutions to realistic problems (related to Data Science) applying the knowledge and skills obtained in different courses, new technologies and current industry practices.
 - a. The institution / Department shall network with the industry to get internships for students.
3. The department will appoint a project coordinator and Internal supervisor who will coordinate the following:
 - Collection of project topics/ descriptions from faculty members (Problems can also be invited from the industries allocating the students to industry for full period of 4-5 months)
 - Allotment of projects and project guides to students
 - Conducting seminar presentations (minimum three)
 - Coordinating viva-voce exam
4. The coordinator shall prepare seminar schedule for all the students (batch wise) from the 5th week to the last week of the semester which should be strictly adhered to. The coordinator will prepare seminar schedule for all the students from the 5th week to the last week of the semester, which should be strictly adhered to.

Each student will be required to:

1. Submit a one-page synopsis before the seminar for display on notice board.
2. Give a 30 minutes presentation followed by 10 minutes discussion.
3. Submit a technical write-up on the talk


Chairperson
B.O.S. in Statistics
U.C.S.
Osmania University
HYDERABAD-7


Professor, Dept. of Statistics
UNIVERSITY COLLEGE OF SCIENCES
Osmania University
HYDERABAD-7, T.S.


HEAD
Department of Computer Science
Telangana Mahila Viswavidyalayam
(Women's University)
Koti, Hyderabad-500 095, T.S.