M.Sc. Data Science Course Curriculum

(w.e.f. Academic Year 2023-24)

M.SC. (Data Science) FIRST YEAR

I-SEMESTER

Paper Code	Title of the paper	Credits	# Hours	Max. Marks
MDS-101	Mathematical Foundations for Data Science	4	4	100
MDS-102	Design and Analysis of Algorithms	4	4	100
MDS-103	Software Engineering	4	4	100
MDS-104	A: Principles of Data Science B: Java Programming	4	4	100
MDS-105	Design and Analysis of Algorithms Lab	2	4	50
MDS-106	A: Principle of Data Science Lab B: Java Programming Lab	2	4	50
	Total	**20**	**24**	**500**

II-SEMESTER

Paper Code	Title of the paper	Credits	# Hours	Max. Marks
MDS-201	Statistical Inference	4	4	100
MDS-202	Data Visualization Techniques	4	4	100
MDS-203	Cloud Computing	4	4	100
MDS-204	Artificial Intelligence	4	4	100
MDS-205	Advanced Machine Learning Techniques	4	4	100
MDS-206	Statistical Inference & Data Visualisation Lab	2	4	50
MDS-207	Advanced Machine Learning Lab	2	4	50
	Total	**24**	28	**500**

M.SC. (Data Science) SECOND YEAR

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 (E-I)	a) Data Mining	4	4	100
	b) Text Data Analytics			
	c) Enterprise Architecture			
	d) Business Intelligence			
MDS-304 (E-II)	a) Data Stream Mining	4	4	100
	b) Sentimental Analysis			
	c) Scalable Architecture			
	d) Computer Vision			
MDS-305	Deep Learning Techniques Lab	2		
MDS-306	Capstone Project-I	2	4	50
MDS-307	Seminar	2	4	50
	No of Credits	**22**	**24**	**500**

Elective Streams E-I & E-II:

A: Data Mining and Data stream mining
B: Text Data Analytics and Sentimental Analysis
C: Architecting Applications and Scalable Architecture
D: Business Intelligence and Computer Vision

IV-SEMESTER

Paper Code	**Title of the paper**	**Credits**	**# Hours**	**Max. Marks**
MDS-401	Research Methodology	2	2	100
MDS-402	Capstone Project-II	12	24	100
	No of Credits	**24**	26	**500**

M.SC. (DATA SCIENCE) I-SEMESTER

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

SYLLABUS

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. (DATASCIENCE) I-SEMESTER

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

SYLLABUS

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.

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. Pannerselvam (2007): Design and Analysis of Algorithms, PHI.
4. Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

M.SC. (DATA SCIENCE) I-SEMESTER

MDS-103: PAPER- III: SOFTWARE ENGINEERING

SYLLABUS

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.

UNIT – IV

Software Configuration Management, SCM Process, Product Metrics for Requirements Model, Design Model, Source Code, Testing and Maintenance. Managing Software Projects: The Project Management Spectrum, W5HH Principle, Metrics in the Process and Project Domains, Software Measurement, Metrics for Software Quality, Integrating Metrics within the Software Process, Software Project Estimation, Decomposition Techniques, Project Scheduling – basics, scheduling, Software Risks, Risk Mitigation, Monitoring, and Management, Software Maintenance, Software Reengineering, Reverse Engineering, Forward Engineering.

REFERENCE BOOKS

1. Roger S Pressman, B R Maxim, Software Engineering – A Practitioner's Approach (8e) References
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

M.SC. (DATASCIENCE) I-SEMESTER

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

(Mandatory for B.Sc. Computer Science Stream)

SYLLABUS

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. (DATASCIENCE) I-SEMESTER

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

(Mandatory for B.Sc. Data Science stream)

SYLLABUS

UNIT – I

Core Java: Class Object, Object Oriented Concepts with respect to Java, Interfaces, Packages and Exception Handling, Applets, Overview of Collection Framework (No question to be set from above topics). AWT: Introduction, AWT Class Hierarchy, Creating Container, Adding Components, Layout, Using Panel, Text Field, Text Area, List, Checkbox, Check Box Group, Choice, Event Handling, Dialog Boxes, ScrollBar, Menu. Swing: Containment Hierarchy, Adding Components, JTextField, JPasswordField, JTable, JComboBox, JProgressBar, JList, JTree, JColorChooser, Dialogs. Remote Method Invocation (RMI): Introduction, Remote Method Invocation, Java RMI Interfaces and Classes, an Application, Compiling the Program, Generating Stub Classes, Running the Program, Callback with an Application.

UNIT – II

Servlet: Server-Side Java, Servlet Alternatives, Servlet Strengths, Servlet Architecture, Servlet Life Cycle, GenericServlet, HttpServlet, Servlet Example, Passing Parameters to Servlets, Retrieving Parameters, Cookies, Filters. Java Server Pages (JSP): Introduction, JSP Engines, How JSP Works, JSP and Servlet, Anatomy of a JSP Page, JSP Syntax, JSP Components, Beans, Session Tracking, Users Passing Control and Data between Pages, Sharing Session and Application Data.

UNIT – III

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, Result Set Metadata. Hibernate: Introduction, Writing POJO Class, Creating a Table, Writing a Hibernate Application, Compiling and Running Application, Book Application Using Annotation, Object Life Cycle, HQL, Using Native SQL Query, Named Queries, Generating DDL, Generator Class, Hibernate Tools.

UNIT – IV

Java Naming and Directory Interface (JNDI): Naming Concepts, Directory Concepts, Java Naming and Directory Interface, Specifying JNDI Properties, Name Servers, Naming Operations, Working with Directory. Overview of J2EE, Introduction to JavaBeans, Advantages of JavaBeans, Properties of JavaBeans with examples, JavaBeans API, Enterprise JavaBeans (EJB), Applications using Session Beans and Entity Beans, Introduction to Struts Framework. Java Server Faces (JSF): Introduction, Simple Application, Request Processing Life-Cycle, Tracing Phases, Managed Bean, Basic JSF Tags, Expression Language, Event Handling with Example, Page Navigation.

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

M.SC. (DATASCIENCE) I-SEMESTER

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

M.SC. (DATASCIENCE) I-SEMESTER

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.

M.SC. (DATASCIENCE) I-SEMESTER

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

List of Experiments

1. 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.
2. Create GUI to demonstrate typical Editable Table which describing Employee for a software company.
3. Create GUI to demonstrate swing components using student registration form.
4. Create a Remote Object for simple arithmetic operators. Use AWT/SWING to create user interface.
5. Write an RMI application using call back mechanism
6. Develop Servlet Question-Answer Application using HttpServletRequest and HttpServletRequest interfaces.
7. Develop Servlet application to accept HTNO of a student from client and display the memorandum of marks from the server
8. 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.
9. Write a simple JSP application to demonstrate the use of implicit object (at least 5).
10. Develop a Hibernate application to Store Feedback of Website Visitors in MySQL Database.
11. Develop a JSP Application to accept Registration Details from the user and store database table.
12. Develop a JSP Application to Authenticate User Login as per the Registration Details. If Login Success then forward User to Index Page otherwise show Login failure Message.
13. Develop a web Application to add items in the inventory using JSF.
14. Write EJB applications using stateless session beans and state-full session beans.
15. Develop a Room Reservation System Application using Entity Beans.
16. Create Three-tire application using Servlets, JSP, EJB.

M.SC. (DATASCIENCE) II-SEMESTER

MDS-201: PAPER- I: STATISTICAL INFERENCE

SYLLABUS

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, Kernal 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. 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)

M.SC. (DATA SCIENCE) II-SEMESTER
MDS-202: PAPER- II: DATA VISUALIZATION TECHNIQUES
SYLLABUS

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 (20 16): The Visual Display of Quantitative Information" by
3. Kieran Healy(20160: Data Visualization: A Practical Introduction,
4. Claus O. Wilke (2020) Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures:, O'relly
5. Stephanie D.H. Evergreen(2020): Effective Data Visualization: The Right Chart for the Right Data

M.SC. (DATASCIENCE) II-SEMESTER

MDS-203: PAPER- III: CLOUD COMPUTING

SYLLABUS

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 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) II-SEMESTER

MDS-204: PAPER- IV: ARTIFICIAL INTELLIGENCE

SYLLABUS

UNIT - I

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.

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.

M.SC. (DATASCIENCE) II-SEMESTER

MDS-205: PAPER- V: ADVANCED MACHINE LEARNING TECHNIQUES

SYLLABUS

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.

M.SC. (DATASCIENCE) II-SEMESTER

MDS-206 (LAB-1): PAPER- VI: 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).

M.SC. (DATASCIENCE) II-SEMESTER

MDS-207: (LAB-2): PAPER- VII: 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.

M.SC. (DATASCIENCE) III-SEMESTER

MDS-301: PAPER- I: DEEP LEARNING TECHNIQUES

SYLLABUS

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.

M.Sc. (DATA SCIENCE) III-SEMESTER

MDS-302: PAPER- II: MACHINE LEARNING OPERATIONS

SYLLABUS

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 BOOKS

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) III-SEMESTER

MDS-303 (E-I): PAPER- III (A): DATA MINING

SYLLABUS

UNIT-I

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

UNIT-II

Mining frequent patterns, Associations and correlations: Basic concepts and methods, Frequent Item set Mining Methods, Which patterns are interesting? Pattern evaluation methods.

UNIT-III

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

UNIT-IV

Cluster Analysis: Concepts and Methods: Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Evaluation of clustering.

UNIT-V

Data Mining Trends and Research Frontiers: Mining Complex Data Types, Other Methodologies of Data Mining, Data Mining Applications, Data Mining and Society, Data Mining trends.

REFERENCE BOOKS

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

M.SC. (DATASCIENCE) III-SEMESTER

MDS-303 (E-I): PAPER- III (B): TEXT DATA ANALYTICS

SYLLABUS

UNIT - I

Introduction to Natural Language Processing Basics, 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

UNIT – IV

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 …), concept of Sentiment Analysis and its applications. Sentence Subjectivity and Sentiment Classification; **Sentiment Lexicon generation and Summarization;**

REFERENCE BOOKS

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

MDS-303 (E-I): PAPER- III (C): ENTERPRISE ARCHITECTURE

SYLLABUS

UNIT-I

Introduction to EA -System analysis, general system theory, definitions and objectives of considerations, Properties of EA, system approach to EA development, principle definitions

Business architecture, definition and features- BSC – balanced score card basics and its reflection in EA, Strategic governance, Event Causality effects in EA under scope of BSC

UNIT-II

Organizational structure of EA and basic models- Information and technology architecture basics, Introduction to EA structuring and modeling, Business architecture (inc. business process modeling, IBM Component business model), Information architecture, Technology architecture and integration between the layers model.

UNIT-III

Introduction to enterprise Engineering (EE)- Enterprise transformations (waterfall and agile), EAP, EA methodologies: PRISM, ARIS Framework, Zachmann Framework , FEAF, DODAF and TOGAF, Introduction to Service orientation in Enterprise Engineering (SOA, SoEA), Technological infrastructure for Big Data handling in EA.

UNIT-IV

Cloud Computing Opportunities for EA- Flexible (agile) business and information architectures (SoEA). Introduction to Spark, Spark Data Frames, SQL, Datasets through worked examples. Spark's low level APIs, RDDS, execution of SQL & Data Frames. How Spark Runs on a Cluster. Structured Streaming, Spark's Stream – Processing Engine.

REFERENCE BOOKS

1. Designing Enterprise Architecture Frameworks: Integrating Business Processes with IT Infrastructure by N Zarvić, R Wieringa. Apple Academic Press (19 April 2016).
2. Neubauer M., Stary CH., S-BPM in the Production Industry. Stakeholder approach, Springer Open, 2017.
3. A systematic literature review on Enterprise Architecture Implementation Methodologies by Babak D., Mohd N. Elsevier (June 2015).
4. Spark : The Definite Guide – Bill Chambers, Matei Zaharia, 2018

M.Sc. (DATA SCIENCE) III-SEMESTER

MDS-303 (E-I): PAPER- III (D): BUSINESS INTELLIGENCE

SYLLABUS

UNIT – I

Business Intelligence: Introduction – Definition, Leveraging Data and Knowledge for BI, BI Components, BI Dimensions, Information Hierarchy, Business Intelligence and Business Analytics. BI Life Cycle. Data for BI - Data Issues and Data Quality for BI.

UNIT - II

BI Implementation - Key Drivers, Key Performance Indicators and Performance Metrics, BI Architecture/Framework, Best Practices, Business Decision Making, Styles of BI-vent-Driven alerts - A cyclic process of Intelligence Creation. The value of Business Intelligence-Value driven & Information use.

UNIT - III

Advanced BI – Big Data and BI, Social Networks, Mobile BI, emerging trends, Description of different BI-Tools (Pentaho, KNIME)

UNIT-IV

Business intelligence implementation-Business Intelligence and integration implementation-connecting in BI systems- Issues of legality- Privacy and ethics- Social networking and BI.

REFERENCE BOOKS

1. Rajiv Sabherwal "Business Intelligence" Wiley Publications, 2012.
2. Efraim Turban, Ramesh Sharda, Jay Aronson, David King, Decision Support and Business Intelligence Systems, 9th Edition, Pearson Education, 2009.
3. David Loshin, Business Intelligence - The Savy Manager's Guide Getting Onboard with Emerging IT, Morgan Kaufmann Publishers, 2009.
4. Philo Janus, Stacia Misner, Building Integrated Business Intelligence Solutions with SQL Server, 2008 R2 & Office 2010, TMH, 2011.
5. Business Intelligence Data Mining and Optimization for decision making [Author: Carlo-Verellis] [Publication: (Wiley)].

M.SC. (DATASCIENCE) III-SEMESTER

MDS-304 (E-II): PAPER- IV (A): DATA STREAM MINING

SYLLABUS

UNIT-I

MOA Stream Mining, Assumptions, Requirements, Mining Strategies, Change Detection Strategies, MOA Experimental Settings, Previous Evaluation Practices, Evaluation Procedures for Data Streams, Testing Framework, Environments, Data Sources, Generation Speed and Data Size, Evolving Stream Experimental Setting.

UNIT-II

Hoeffding Trees, The Hoeffding Bound for Tree Induction, The Basic Algorithm, Memory Management, Numeric Attributes, Batch Setting Approaches, Data Stream Approaches.

UNIT-III

Prediction Strategies, Majority Class, Naïve Bayes Leaves, Adaptive Hybrid, Hoeffding Tree Ensembles, Data Stream Setting, Realistic Ensemble Sizes. Evolving Data Streams, Algorithms for Mining with Change, A Methodology for Adaptive Stream Mining, Optimal Change Detector and Predictor, Adaptive Sliding Windows, Introduction, Maintaining Updated Windows of Varying Length.

UNIT-IV

Adaptive Hoeffding Trees, Introduction, Decision Trees on Sliding Windows, Hoeffding Adaptive Trees, Adaptive Ensemble Methods, New methods of Bagging using trees of different size, New method of bagging using ADWIN, Adaptive Hoeffding Option Trees, Method performance.

REFERENCE BOOKS

1. Data Stream Mining: A Practical Approach by Albert Bifet and Richard Kirkby.
2. Knowledge discovery from data streams by Gama João. ISBN: 978-1-4398-2611-9.
3. Machine Learning for Data Streams by Albert Bifet, Ricard Gavalda; MIT Press, 2017.

M.Sc. (DATA SCIENCE) III-SEMESTER

MDS-304 (E-II): PAPER- IV (B): SENTIMENT ANALYSIS

SYLLABUS

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.

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 BOOK

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

M.Sc. (DATA SCIENCE) III-SEMESTER

MDS-304 (E-II): PAPER- IV (C): SCALABLE ARCHETECTURE

SYLLABUS

UNIT–I

Introduction to Scalable applications, Challenges with running applications using Machine Learning with scaling, Algorithms for Large scale Learning, Overview of Hadoop and Current Big Data Systems. How Programming for Data Flow Differs, Basic Spark, Working with Vectors and Matrices in Spark, Brief tour of Spark ML, Beyond parallelization, Practical Big Data.

UNIT–II

Anatomy of Fast Data Applications, SMACK Stack – Functional Decomposition, Message Backbone- Understanding messaging requirements, Data ingestion, Fast data& low latency, Message Delivery Semantics, Distributing Messages.

UNIT – III

Compute Engines- Micro Batch Processing, One-at-a time Processing, Choice of processing engine, Storage as the Fast Data Borders, The message backbone as Transition Point

UNIT-IV

Sharing Stateful Streaming State, Data Driven Micro-services, State and Micro-services. Deployment environments for Fast Data Applications, Application containerization, resource scheduling, Apache Mesos, Kubernetes, Cloud Deployments.

REFERENCE BOOKS

1. Jan Kunigk, Ian Buss, Paul Wilkinson & Lars George," Architecting Modern 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. (DATASCIENCE) III-SEMESTER

MDS-304 (E-II): PAPER- IV (D): COMPUTER VISION

SYLLABUS

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 matching: Points and patches – Edges – Lines. Segmentation: Active Contours – Split and merge – Mean shift and mode finding – Normalized cuts –Graph cuts and energy –based methods

UNIT-IV

Recognition: Object detection – Face recognition – Instance recognition – Category Recognition – Context and scene understanding –Recognition datasets and test sets

REFERENCE BOOKS

1. Richard Szeliski (2011): Computer Vision-Algorithms and Applications, Springer verlang London Limited.
2. Deep Learning, by Goodfellow, Bengio, and Courville.
3. Dictionary of Computer Vision and Image Processing, by Fisher et al.

M.SC. (DATASCIENCE) III-SEMESTER

MDS-305 (LAB-I): PAPER- V: DEEP LEARNING TECHNIQUES LAB

SYLLABUS

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.

M.SC. (DATASCIENCE) III-SEMESTER

MDS-306 (LAB-2): PAPER- VI: CAPSTONE PROJECT PHASE -I

Project is of two-semesters duration (III and IV Semesters). Each student carries out a project individually. A student with the help of the guide formulates the problem during the III semester and implements it during the IV semester. A student can carry out the project at industry for the entire IV Semester.

Note: Objectives, Outcomes and Guidelines are given under Capstone Project Phase-II.

M.SC. (DATASCIENCE) III-SEMESTER

MDS-308 (LAB-3): PAPER- V: SEMINAR
(Guidelines For Seminar)

Objectives:

1. To familiarize tools and techniques and content for presentation
2. To enhance practical presentation, effective communication and professional skills
3. To expose the students to answer the queries raised on the topic of presentation.
4. To encourage students to work with innovative and entrepreneurial ideas

The objective of the seminar is to prepare the student for a systematic and independent study of state-of-the-art topics in a broad area of his / her specialization. Seminar topics may be chosen by the students with advice from the faculty members. Students are to be exposed to following aspects of seminar presentations; Literature survey, Organization of material, power point presentation, technical writing Each student will be required to

1. Each student has to give minimum two seminars.
2. Submit a one page synopsis of the seminar talk for display on the notice board.
3. Give a 20 minutes presentation through power point presentation followed by 10 minutes of discussion.
4. Submit a report on the seminar topic with list of references.

Seminars are to be scheduled from the 2nd week to the last week of the semester and any change in schedule should be discouraged. The sessional marks will be awarded to the students by at least 2 faculty members on the basis of an oral and a written presentation as well as their involvement in the discussions.

M.SC. (DATA SCIENCE) IV-SEMESTER

MDS/401: PAPER- I: RESEARCH METHODOLOGY

SYLLABUS

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

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 - II

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.

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 BOOKS

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

M.SC. (DATA SCIENCE) IV-SEMESTER

MDS-402: PAPER- II: CAPSTONE PROJECT PHASE -II

(Guidelines continuation of Phase-I of III-Semester)

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 theacademic 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

Project is of 2-semester duration (III and IV Semesters). Each student carries out a project individually. A student with the help of the guide formulates the problem during the III semester and implements it during the IV semester. A student can carry out the project at industry for the entire IV Semester. 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.

1. The institution / Department shall network with the industry to get internships for students.
2. 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 beinvited 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

   The above tasks should be completed within the first two weeks of III semester.

To get awareness on current problems and solution techniques, Project coordinator shall arrange special lectures during the first 2 weeks of III semester by inviting faculty

members, professionals from industries and R&D institutions. Further, these lectures may be conducted anytime during the semester to enable the students to gather information on problems and industry practices. At the end of 2nd week, each student with the help of guide shall formalise the project proposal with problem definition, scope, literature survey, probable solution etc. 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 lastweek 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.