.

Rayat Shikshan Sanstha's

SADGURU GADGE MAHARAJ COLLEGE, KARAD

(An Autonomous College)

Re accredited by NAAC with 'A+' Grade.

Syllabus for Master of Science

Syllabus

M.Sc. Data Science

Two Year Degree Program in Data Science

(Faculty of Science)

As per NEP-2020

To be implemented from academic year 2024-25

Title of the Course: M.Sc. (Data Science)

Preamble:

In today’s tech-driven world, access to vast amounts of information and ways to interpret it have taken priority than ever before. Real time processing of this huge data is also a major requirement in every walk of life. It also means we need more people who can organize and analyze that information - people who can use data to make change and help businesses. Data science employs a variety of instruments, scientific procedures, methods, and algorithms to glean insights from both structured and unstructured data. This Data Science program integrates scientific methods from statistics, computer science and data-based business management to extract knowledge from data and drive decision making. Our curriculum provides students with a rigorous course of study in big data technologies, applications and practices a pathway for student internships and full-time employment. Students are prepared to meet the challenges at the intersection between big data, business analytics, and other emerging fields.

Eligibility:

- B.Sc.(Statistics / Mathematics / Electronics / Physics / Chemistry) from recognized Indian University.
- B.Sc.(Computer Science), B.C.S., B.Sc.(Entire Computer Science), (BCA)Bachelor of Computer Application from recognized Indian University.

Programme Outcomes (POs)

On completion of M.Sc. Data Science program,students will be able

1. To apply ethical practices in everyday business activities and make well-reasoned ethical business and data management decisions.
2. To demonstrate knowledge of statistical data analysis techniques utilized in business decision making.
3. To apply principles of Data Science to the analysis of business problems.
4. To use data mining software to solve real-world problems.
5. To employ cutting edge tools and technologies to analyze Big Data.
6. To apply algorithms to build machine intelligence.
7. To demonstrate use of team work, leadership skills, decision making and organization theory.

Duration:

- The course shall be a full-time course.
- The course shall be for two years, consisting of four semesters.

Fee Structure

- **Entrance Examination fees:** as prescribed by the Institute.
- **Course Fee:** as prescribed by the Institute.

Medium of instruction: English

- **Structure of the Programme, Scheme of Teaching and Examination**

Sem.	Course type	Course code	No. of credits	Teaching hours per week	Examination Scheme				
					University Assessment			Internal Assessment	
					Max Marks	Min Marks for passing	Exam hours	Max Marks	Min Marks for passing
I	Major Mandatory	MJ-MDS24-101	4	4	80	32	3	20	8
		MJ-MDS24-102	4	4	80	32	3	20	8
		MJ-MDS24-103	2	2	40	16	1.5	10	4
		MJ-MDSP24-106	4	12	--	--	--	100	40
	Major Elective	GE-MDS24-104	4	4	80	32	3	20	8
	Research Methodology	RM-MDS24-105	4	4	80	32	3	20	8
II	Major Mandatory	MJ-MDS24-201	4	4	80	32	3	20	8
		MJ-MDS24-202	4	4	80	32	3	20	8

		MJ-MDS24-203	2	2	40	16	1.5	10	4
		MJ-MDSP24-206	4	12	--	--	--	100	40
	Major Elective	GE-MDS24-204	4	4	80	32	3	20	8
	On Job Training/ Field Project	FP-MDS24-205	4	4	- - 100 40				
III	Major Mandatory	MJ-MDS24-301	4	4	80	32	3	20	8
		MJ-MDS24-302	4	4	80	32	3	20	8
		MJ-MDS24-303	2	2	40	16	1.5	10	4
		MJ-MDSP24-306	4	12	--	--	--	100	40
	Major Elective	GE-MDS24-304	4	4	80	32	3	20	8
	Research Project	RP1-MDS24-305	4	4	--	--	--	100	40
IV	Major Mandatory	MJ-MDS24-401	4	4	80	32	3	20	8

		MJ-MDS24-402	4	4	80	32	3	20	8
		MJ-MDSP24-405	4	12	--	--	--	100	40
	Major Elective	GE-MDS24-403	4	4	80	32	3	20	8
	Research Project	RP2-MDS24-404	6	6	--	--	--	150	60

- **Courses being offered**

SEMESTER – I

Course type	Course code	No. of credits	Course title
Major Mandatory	MJ-MDS24-101	4	Statistics for Data Science
	MJ-MDS24-102	4	Computational Mathematics
	MJ-MDS24-103	2	Fundamentals of Data Science
	MJ-MDSP24-106	4	Practical-I
MajorElective	GE-MDS24-104	4	Python for Data Science
	GE-MDS24-104		Artificial Intelligence
	GE-MDS24-104		Stochastic Models and Applications
Research Methodology	RM-MDS24-105	4	Research Methodology

SEMESTER – II

Course type	Course code	No. of credits	Course title
Major Mandatory	MJ-MDS24-201	4	Data Mining
	MJ-MDS24-202	4	Data Analysis using R
	MJ-MDS24-203	2	Time Series Analysis and Forecasting
	MJ-MDSP24-206	4	Practical-II
Major Elective	GE-MDS24-204	4	Data Base Management System
	GE-MDS24-204		Exploratory Multivariate Data Analysis
	GE-MDS24-204		Baysian Inference
On Job Training/ Field Project	FP-MDS24-205	4	On Job Training industry / Field Projects

Guidelines for On Job Training (FP-MDS24-205)

- Student must start the OJT immediately after semester-II examination during the summer vacation
- Student is expected to complete the field project/OJT work minimum in between 120 hours assigned by company/ industry/ consultancy/ institution.
- College should assign the mentor to monitor the progress throughout the OJT
- Student has to submit the weekly progress report duly signed by the concern authorities of company to the mentor
- At the end of OJT, student should prepare documentation and submit a report.
- The final presentation and documentation will be evaluated by the examination panel.

SEMESTER – III

Course type	Course code	No. of credits	Course title
Major Mandatory	MJ-MDS24-301	4	Machine Learning
	MJ-MDS24-302	4	Regression and Predictive Modeling
	MJ-MDS24-303	2	Big Data
	MJ-MDSP24-306	4	Practical-III
Major Elective	GE-MDS24-304	4	Natural Language Processing
	GE-MDS24-304		Cyber Security
	GE-MDS24-304		Digital Image Processing
Research Project	RP1-MDS24-305	4	Research Project work

Guidelines for Research Project (RP1-MDS24-305)

Student is expected to do the research project work under the guidance of mentor assigned and to present presentation of the research work evaluated by the institute examination panel.

SEMESTER – IV

Course type	Course code	No. of credits	Course title
Major Mandatory	MJ-MDS24-401	4	Deep Learning
	MJ-MDS24-402	4	Data Visualization
	MJ-MDSP24-405	4	Practical IV
Major Elective	GE-MDS24-403	4	Blockchain Technology & Applications
	GE-MDS24-403		Cloud Computing
	GE-MDS24-403		Business Intelligence and Data Analytics
Research Project	RP2-MDS24-404	6	Research Work

<u>Detailed Guidelines of Research Project work for Semester III & IV :</u>

Research Project

Research Project will consist of 2 parts :

I. The Research Proposal

II. The actual dissertation or Research Project Report

I.The Research Proposal :

Students are required to submit their research ideas in the form of a research proposal to their supervisors / advisors / guides and get approval from the guide before the actual research work starts.

❖ **Format of Research Proposal (RP) :**

Project Title

Introduction and Origin of the research problem

Interdisciplinary relevance

Review of Research and Development in the Subject

National / International status

Significance of the study - Objective, methodology

Approximate time by which each stage will be completed Expected results and the outcome(s) of the research project. Bibliography

❖ **Following can be used as a guide to evaluate a RP :**

- Does the proposal address a well-formulated problem? Have research gaps been identified.
- Is it a research problem, or is it just a routine application of known techniques?
- Do the proposers have a good idea on which to base their work? The proposal must explain the idea in sufficient detail to convince the reader that the idea has some substance, and should explain why there is reason to believe that it is indeed a good idea.
- Does the proposal explain clearly what work will be done? Does it explain what results are expected and how they will be evaluated? How would it be possible to judge whether the work was successful?

- Is there evidence that the proposers know about the work that others have done on the problem? This evidence may take the form of Literature Review or a short review as well as representative references.

The proposal should answer three key questions:

a)What are we going to learn as the result of the proposed project that we do not know now?

b)Why is it worth knowing?

c)How will we know that the conclusions are valid?

II)The Research Project

Students should submit a proper research dissertation at the end of their research work for the required credits.

Format of Research Project

Format of Research Project

a) Title of Research
b) Certificate
c) Index
d) List of Figures
e) List of Tables
f) Publications
g) Introduction - Objectives of the Research
h) Literature Review of previous research in the area and justification / Importance /Value of further research, Data, Scope and Limitations
i) Actual Work Done with Experimental Setup, if any.
j) Results and Discussion
k) Future scope of research
l) Bibliography in format –Author name, title, publication details, year

- Standard of passing: 40% in each course. Separate passing for internal and semester examinations.
- Nature of Question paper and Scheme of marking for University examination.

Nature of the theory question papers (4credits):

a) There shall be 7 questions each carrying16 marks.

b) Question No.1i s compulsory . It consists of 8 questions for 2 marks each.

c) Students have to attempt any 4 questions from question No. 2 to 7.

d) Question No. 2to 6 shall contain 2 to 4 sub-questions.

e) Question No.7 shall contain 4 short note type questions ,each carrying4 marks.

Nature of the theory question papers (2credits):

a) There shall be 4 questions.

b) Question No.1 is compulsory. It consists of 4 questions for 2 marks each.

c) Question No. 2 to 4 shall be of 16 marks each.

d) Students have to attempt any 2 questions from question No. 2 to 4.

e) Question No. 2 to 4 shall contain 2 to 4 sub-questions.

- **Nature of Practical examination: -**

Component	**Max marks**
Practical examination: Examination will be of 3hours duration. There shall be 8 questions each of 12marks, of which a student has to attempt any 5 questions.	60
Day-to-day practical performance and journal	20
Viva: Viva will be based on all practical’s	20

SEMESTER I

Course Code: MJ-MDS24-101

Course Title: Statistics for Data Science

Course Outcomes

On completion of the course, students will be able to:

- CO1: Develop relevant programming abilities.
- CO2: Execute statistical analyses with professional statistical software.
- CO3: Demonstrate skill in data management.
- CO4: Develop the ability to build and assess data-based models.

Unit 1: Basics of Statistics & Measures of Central Tendency and Dispersion

Introduction to Statistics, Collection and Scrutiny of Data, Classification and Tabulation of Data, Diagrammatic Presentation of Data, Graphical Presentation of Data, Measures of Central Tendency – Mean, Median and Mode, Measures of Dispersion – Variance, Standard Deviation, Coefficient of Variation, Skewness, and Kurtosis.

Unit 2: Correlation and Linear Regression

Bivariate Data, Scatter Diagram, Correlation – Positive, Negative, and Zero Correlation, Karl Pearson's Coefficient of Correlation (r), Limits of r ($-1 \leq r \leq 1$), Interpretation of r, Coefficient of Determination (r^2), Meaning of Regression and Difference between Correlation and Regression, Fitting of Line $Y = a + bX$, Concept of Residual Plot and Mean Residual Sum of Squares, Multiple Correlation Coefficient – Concept, Definition, Computation, and Interpretation, Partial Correlation Coefficient – Concept, Definition, Computation, and Interpretation.

Unit 3: Probability and Functions of Random Variables

Introduction to Probability, Random Experiments and Empirical Basis of Probability, Algebra of Events and Laws of Probability, Conditional Probability, Independence, and Bayes' Law, Applications of Probability to Business and Economics, Random Variables – One-dimensional (Discrete and Continuous), Distribution Functions and Properties, Bivariate Random Variables – Joint Probability Functions, Marginal Distributions, Conditional Distribution Functions, Notion of Independence, Functions of Random Variables – Distribution Function Technique, Transformation Technique (One and Several Variables), Theory and Applications, Mathematical Expectation, Variance, and Co-variance of Random Variables, Conditional Expectation and Conditional Variance, Chebyshev's Inequality, Weak Law of Large Numbers, Strong Law of Large Numbers, and Central Limit Theorem.

Unit 4: Discrete & Continuous Distributions

Discrete Distributions – Bernoulli, Binomial, Poisson, Geometric, Hypergeometric, Negative Binomial, Multinomial, and Discrete Uniform Distribution – Definition, Properties, and Applications with Numerical Problems; Continuous Distributions – Uniform, Normal, Exponential, Gamma, Beta (First and Second Kind), Weibull, Cauchy, Laplace, Lognormal, Logistic, Pareto, Chi-square, and Rayleigh Distributions – Definition, Properties, and Applications; Concept of Truncated Distributions.

References:

1. Parimal Mukhopadhyay (2012). *An Introduction to the Theory of Probability.* World Scientific.
2. Irwin Miller, Marylees Miller, John E. Freund (2017). *Mathematical Statistics.* Pearson.
3. Fetsje Bijma, Marianne Jonker, Aadvander Vaart (2018). *Introduction to Mathematical Statistics.* Amsterdam University Press.
4. Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Chapman & Hall/CRC.
5. Rohatgi, V.K., Ebsanes Saleh, A.K. Md. (2002). *An Introduction to Probability and Statistics* (2nd Ed.). John Wiley & Sons.
6. Shanmugam, R., Chattamvelli, R. (2015). *Statistics for Scientists and Engineers.* John Wiley.

Course Code: MJ-MDS24-102

Course Title: Computational Mathematics

Course Outcomes

On completion of the course, students will be able to:
CO1 – Effectively use matrix algebra tools to analyze and solve systems of linear equations.
CO2 – Apply numerical methods to solve linear systems of equations.
CO3 – Work on vector maps.
CO4 – Understand the application of mathematics in data science.

Unit 1: Vector Spaces and Matrix Algebra

Vector space, subspace, linear dependence and independence, basis, dimension of a vector space, examples of vector spaces, Gram-Schmidt orthogonalisation process, orthonormal basis, orthogonal projection of a vector, linear transformations, algebra of matrices, types of matrices, row and column spaces of a matrix, elementary operations and elementary matrices, rank and inverse of a matrix.

Unit 2: Matrix Operations and System of Equations

Permutation matrix, reducible and irreducible matrix, Kronecker product, generalized inverse, Moore-Penrose generalized inverse, and solution of a system of homogeneous and non-homogeneous linear equations.

Unit 3: Eigenvalues, Eigenvectors, and the Cayley-Hamilton Theorem

Characteristic roots and vectors of a matrix, algebraic and geometric multiplicities of a characteristic root, right and left characteristic vectors, Cayley-Hamilton Theorem and its applications.

Unit 4: Matrix Decompositions and Applications in Data Science

Spectral decomposition of a real symmetric matrix, singular value decomposition, Cholesky decomposition, real quadratic forms, reduction and classification, index and signature, extrema of a quadratic form, simultaneous reduction of two quadratic forms. Handwritten digits recognition using simple algorithms – classification of handwritten digits using SVD bases and tangent distance, text mining using latent semantic index, clustering, non-negative matrix factorization, and LGK bidiagonalization.

References:

1. Hadley, G. (1962). *Linear Algebra.* Narosa Publishing House.
2. Kenneth H. Rosen (2011). *Discrete Mathematics and Its Applications.* Tata McGraw Hill, 7th Edition.
3. K. Hoffman and R. Kunze (2005). *Linear Algebra.* Pearson Education, 2nd Edition.

Course Code: MJ-MDS24-103

Course Title: Fundamentals of Data Science

Course Outcomes

On completion of the course, students will be able to:
CO1 – Understand basic concepts of Data Science.
CO2 – Perform Exploratory Data Analysis.
CO3 – Obtain, clean/process, and transform data.
CO4 – Detect and diagnose common data issues, such as missing values, special values, outliers, inconsistencies, and localization.
CO5 – Present results using data visualization techniques.

Unit 1: Introduction to Data Science

Introduction to data science, the three V's: Volume, Velocity, Variety, why learn Data Science, applications of Data Science, the Data Science Lifecycle, Data Scientist's Toolbox, types of data: structured, semi-structured, unstructured data, problems with unstructured data, data sources: open data, social media data, multimodal data, standard datasets, measuring data similarity and dissimilarity, data matrix versus dissimilarity matrix, proximity measures for nominal attributes, proximity measures for binary attributes, dissimilarity of numeric data using Euclidean, Manhattan, and Minkowski distances, proximity measures for ordinal attributes.

Unit 2: Data Preprocessing and Visualization

Data objects and attribute types, what is an attribute, nominal, binary, ordinal attributes, numeric attributes, discrete versus continuous attributes, data quality, why preprocess the data, data munging/wrangling operations, cleaning data including missing values, noisy data problems such as duplicate entries, multiple entries for a single entity, missing entries, nulls, huge outliers, out-of-date data, artificial entries, irregular spacings, formatting issues across different tables/columns, extra whitespace, irregular capitalization, inconsistent delimiters, irregular null formats, invalid characters, incompatible date-times, introduction to exploratory data analysis, data visualization and visual encoding, data visualization libraries, donut charts, specialized data visualization tools including boxplots, bubble plots, heat maps, dendrograms, Venn diagrams, tree maps, 3D scatter plots.

References:

1. Jiawei Han and Micheline Kamber, *Data Mining Concepts and Techniques*, Elsevier.
2. Margaret H. Dunham, *Data Mining: Introductory and Advanced Topics*, Pearson Education.
3. Pang Ning Tan, Michael Steinbach, Vipin Kumar, *Introduction to Data Mining*, Pearson Education.
4. Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning Vol. 1*, Cambridge: MIT Press.
5. S.M. Ross, *Introduction to Probability Models*, Academic Press.

Course Code: GE-MDS24-104
Course Title: Python for Data Science

Course Outcomes

On completion of the course, students will be able to:

CO1 – Manipulate and process datasets.

CO2 – Perform data analysis to find hidden patterns from datasets.

CO3 – Visualize datasets in terms of different charts.

CO4 – Determine methods to create and develop Python programs by utilizing data structures like lists, dictionaries, tuples, and sets.

CO5 – Develop Python programs and create a small application project.

Unit 1: Introduction to Python

The Python programming language, history, features, applications, installing Python, running simple Python programs, comments, data types, variables, operators and operator precedence, data type conversions, simple input and output, command line arguments and data input, introduction to Python IDEs such as PyCharm, Jupyter, and Spyder, conditional statements including if, if-else, nested if-else, looping constructs including for, while, nested loops, loop control statements (break, continue, pass), strings including declaration, manipulation, special operations, escape characters, string formatting operators, raw strings, Unicode strings, built-in string methods.

Unit 2: Lists, Functions, Tuples, Dictionaries, and Sets

Python lists including concept, creating and accessing elements, updating and deleting lists, traversing a list, reverse, built-in list operators, concatenation, repetition, in operator, built-in list functions and methods, functions including definition and use, function calls, type conversion functions, math functions, composition, adding new functions, flow of execution, parameters and arguments, variables and parameters, stack diagrams, void functions, anonymous functions, importing with from, return values, Boolean functions, recursion, functional programming tools including filter(), map(), reduce(), tuples including accessing values, tuple assignment, tuples as return values, variable-length argument tuples, basic tuple operations, concatenation, repetition, in operator, iteration, built-in tuple functions, indexing, slicing, matrices, dictionaries including creating a dictionary, accessing values, updating, deleting elements, properties of dictionary keys, operations in dictionary, built-in dictionary functions and methods, sets including definition, operations such as adding, union, intersection, and working with sets.

Unit 3: Python Modules, Working with Files, Exception Handling

Modules including importing modules, creating and exploring modules, math module, random module, time module, packages including importing and creating packages with examples, working with files including creating files, operations on files (open, close, read, write), file object attributes, file positions, listing files in a directory, testing file types, removing files and directories, copying and renaming files, splitting pathnames, creating and moving directories, exception handling including built-in exceptions, handling exceptions, exceptions with arguments, and user-defined exceptions.

Unit 4: Working with NumPy and Pandas

Installing and launching Jupyter , installing NumPy, NumPy introduction, NumPy datatypes, NumPy arrays, NumPy arithmetic operations, binary operators, NumPy string functions, mathematical functions, statistical functions, NumPy sort, search, and counting functions, installing Spyder and Pandas, introduction to Pandas, Pandas DataFrame object, importing data (.csv, .xlsx, .txt) into Spyder, attributes of data, creating copies of original data, data preprocessing including indexing and selection, handling missing data, operations on null values, frequency tables, two-way tables including joint probability, marginal probability, conditional probability, aggregation and grouping, simple aggregation in Pandas, group by operations including split, apply, combine.

References:

1. Bill Lubanovic, *Introducing Python*
2. John Paul Mueller, *Machine Learning (in Python and R) For Dummies*
3. Dr. R. Nageswara Rao, *Core Python Programming*
4. David Beazley and Brian K. Jones, *Python Cookbook*
5. David Ascher and Alex Martelli, *Python Cookbook*

Course Code: GE-MDS24-104

Course Title: Artificial Intelligence

Course Outcomes

On completion of the course, students will be able to:

CO1 – Learn the basics of AI.

CO2 – Understand the informed and uninformed problem types and apply search strategies to solve them.

CO3 – Apply difficult real-life problems in a state space representation and solve them using AI techniques such as searching and game playing.

CO4 – Design and evaluate intelligent expert models for perception and prediction from intelligent environments.

CO5 – Examine the issues involved in knowledge bases, reasoning systems, and planning.

Unit 1: Introduction to Artificial Intelligence

What is Artificial Intelligence, forms of AI, purpose of AI, applications of AI, what is data science, artificial intelligence in data science, role of artificial intelligence in data science, comparison of AI and data science.

Unit 2: Intelligent Systems

What is an intelligent agent in AI, types of intelligent agents, structure of intelligent agents, properties of intelligent agents, examples of intelligent agents, AI problems (state space search), water jug problem, 8-puzzle problem, travelling salesman problem, Tower of Hanoi problem.

Unit 3: Search Algorithms in AI

Problem solving agents, search algorithm terminologies, properties of search algorithms, types of search algorithms, uninformed/blind search including BFS (Breadth First Search), DFS (Depth First Search), DLS (Depth Limited Search), IDDFS (Iterative Deepening DFS), UCS (Uniform Cost Search), BS (Bi-Directional Search), informed search including best-first search algorithm (Greedy Search), A* search algorithm, AO* search algorithm.

Unit 4: Knowledge Representation

Knowledge-based agent architecture, inference system, what is knowledge representation, types of knowledge, AI knowledge cycle, techniques of knowledge representation, logical representation, semantic network representation, frame representation, production rules, propositional logic, predicate logic (first-order logic), forward and backward chaining, knowledge representation structures including weak structures, strong structures, semantic networks, frames, conceptual dependencies, scripts.

References:

1. Elaine Rich and Kelvin Knight, *Artificial Intelligence*, Tata McGraw Hill.
2. Nils J. Nilson, *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann Publishers, Inc., San Francisco, 2000.
3. Saroj Kaushik, *Artificial Intelligence*, Cengage Learning; B. Yegnanarayana, *Artificial Neural Networks*, Prentice-Hall of India.
4. G. Rajasekaran and A. Vijayalakshmi Pai, *Artificial Intelligence*, Prentice-Hall of India, 2003.
5. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 2nd Edition, Prentice Hall.

Course Code: GE-MDS24-104

Course Title: Stochastic Models and Applications

Course Outcomes

On completion of the course, students will be able to:

CO1 – Formulate TPM (transition probability matrix) and n-step transition probabilities.

CO2 – Classify states in stochastic processes.

CO3 – Perform stochastic simulations.

CO4 – Become familiar with stochastic processes, including Poisson process, Wiener process, and Renewal process.

Unit 1: Notion of Stochastic Processes and Markov Chains

Notion of stochastic processes, Markov chain, one-step transition probabilities, Chapman-Kolmogorov equations, evaluation of higher-step transition probabilities, classification of states, periodicity of a Markov chain, concept of closed class, minimal closed class, stationary distribution, examples such as gambler’s ruin problem and one-dimensional random walk, concept of absorption probabilities, using these to compute the probability of winning the game by a gambler having initial capital ‘a’.

Unit 2: Branching Processes and Continuous-Time Markov Chains

Branching process, classification of states, identification of criticality parameter, extinction probability, relationship between criticality parameter and extinction probability of the process, expression for mean and variance of the process, some epidemiological applications, introduction to Markov chain in continuous time, concept of intensity rate, relationship between intensity matrix and transition probability matrix, Kolmogorov’s forward and backward equations.

Unit 3: Birth and Death Processes

Introduction to birth process, birth and death process, linear birth and death process, growth model with immigration and related results, expression for mean and variance of a birth process and birth and death process, applications of these processes.

Unit 4: Poisson and Renewal Processes

Poisson process, two definitions and their equivalence, distribution of inter-arrival times, conditional joint distribution of inter-arrival times, compound Poisson process, some applications, introduction to renewal process, relationship with Poisson process, key and elementary renewal theorems associated with renewal processes.

References:

1. Bhat, B.R. (2000). *Stochastic Models: Analysis and Applications*, New Age International.
2. Medhi, J. (2010). *Stochastic Processes*, New Age Science Ltd.
3. Pinsky, M.A. and Karlin, S. (2010). *An Introduction to Stochastic Modeling*, 4th Edn., Academic Press.
4. Ross, S. (2014). *Introduction to Probability Models*, 11th Edn., Academic Press.
5. Feller, W. (1972). *An Introduction to Probability Theory and its Applications, Vol. 1*, Wiley Eastern.
6. Hoel, P.G., Port, S.C., Stone, C.J. (1972). *Introduction to Stochastic Processes*, Houghton Mifflin.
7. Karlin, S., Taylor, H.M. (1975). *A First Course in Stochastic Processes*, 2nd Edition, Academic Press.
8. Serfozo, R. (2009). *Basics of Applied Stochastic Processes*, Springer.

Course Code: RM-MDS24-105

Course Title: Research Methodology

Course Outcomes

On completion of the course, students will be able to:

CO1 – Understand the concept of research, research process, and research ethics.

CO2 – Understand and apply various sampling methods for data collection and estimate parameters.

CO3 – Understand the concept of simulation and simulate real-life processes.

CO4 – Estimate bias and standard error of an estimator using resampling techniques, apply numerical methods to solve systems of linear equations, obtain roots of a nonlinear equation, and solve definite integrals.

Unit 1: Concept of Research and Research Ethics

Meaning of research, objectives of research, motivation in research, types of research, research approaches, significance of research, research methods vs. methodology, research and scientific method, research process, criteria of good research, defining research problem, research design, research ethics, publication of research, plagiarism, intellectual property rights, patents and its filing procedures.

Unit 2: Sampling Techniques and Estimation

Review of simple random sampling, stratified random sampling, systematic random sampling, cluster sampling, two-phase sampling, ratio and regression method of estimation, probability proportional to size sampling including cumulative total method, Lahiri's method, Hansen-Horwitz estimator and its properties, Horwitz-Thompson estimator, Des Raj estimators for general sample size, non-sampling errors, techniques for handling non-response including Hansen-Horwitz and Deming's model for effect of callbacks, randomized response techniques, dichotomous population, Warner's model, MLE in Warner's model, unrelated question model.

Unit 3: Interpretation and Report Writing

Meaning of interpretation, why interpretation, technique of interpretation, precautions in interpretation, significance of report writing, different steps in writing report, layout of the research report, types of reports including research proposal/synopsis, research paper, and thesis, oral presentation, mechanics of writing a research report, precautions for writing research reports.

Unit 4: Publication Ethics and Open Access Publishing

Publication ethics including definition, introduction, and importance, best practices/standards setting initiatives and guidelines such as COPE and WAME, conflicts of interest, publication misconduct including definition, concept, problems that lead to unethical behavior and vice versa, types, violation of publication ethics, authorship and contributorship , identification of publication misconduct, complaints and appeals, predatory publishers and journals, open access publications and initiatives, SHERPA/RoMEO online resource to check publisher copyright and self-archiving policies, software tools to identify predatory publications developed by SPPU, journal finder/journal suggestion tools including JANE, Elsevier Journal Finder, Springer Journal Suggester, E-resources for research including Google Scholar, ShodhGanga, ShodhGangotri, SciHub.

References:

1. Atkinson, K.E. (1989). *An Introduction to Numerical Analysis*, John Wiley and Sons.
2. Chaudhuri, A., & Stenger, H. (2005). *Survey Sampling: Theory and Methods*, CRC Press.
3. Cochran, W.G. (1977). *Sampling Techniques*, John Wiley & Sons.
4. Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
5. Efron, B., & Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*, CRC Press.
6. Kennedy, W.J., & Gentle, J.E. (2021). *Statistical Computing*, Routledge.
7. Kothari, C.R. (2004). *Research Methodology: Methods and Techniques*, New Age International.
8. Morgan, B.J. (1984). *Elements of Simulation (Vol. 4)*, CRC Press.
9. Mukhopadhyay, P. (2008). *Theory and Methods of Survey Sampling*, PHI Learning Pvt. Ltd.
10. Robert, C.P., Casella, G., & Casella, G. (1999). *Monte Carlo Statistical Methods (Vol. 2)*, Springer, New York.
11. Ross, S.M. (2022). *Simulation*, Academic Press.
12. Rubinstein, R.Y., & Melamed, B. (1998). *Modern Simulation and Modeling (Vol. 7)*, Wiley, New York.
13. Singh, D., & Chaudhary, F.S. (1986). *Theory and Analysis of Sample Survey Designs*, John Wiley & Sons.
14. Sukhatme, P.V., Sukhatme, S., & Ashok, C. (1984). *Sampling Theory of Surveys and Applications*, Iowa University Press and Indian Society of Agricultural Statistics, New Delhi.

Course Code: MJ-MDSP24-106

Course Title: Practical-I

1. Diagrammatic Representation and Descriptive Statistics for raw data (examples: bar chart, line chart, pie chart, etc.)
2. Looking at the data: Data Summaries: Measures of Central Tendency, Measures of Dispersion, Measures of Skewness and Kurtosis.
3. Implementation of Correlation and Linear Regression
4. Case study (Using real-world dataset, for example, Kaggle dataset – https://www.kaggle.com), students are supposed to perform all above experiments for statistical analysis of data.
5. Formulation of research problem and its design.
6. Sampling techniques-I
7. Sampling techniques-II
8. Applications of Simulation techniques.
9. Numerical Methods and Resampling Techniques.
10. A Mini Case Study on the Impact of Social Media Usage on the Academic Performance of College Students.
11. Practical based on application of vector spaces.
12. Practical based on Matrix Operations and System of Equations
13. Practical based on Eigenvalues, Eigenvectors and Cayley-Hamilton Theorem.
14. Practical based on Matrix Decompositions and Applications in Data Science.
15. Write a program to create, concatenate, and print a string and access substring from a given string.
16. Write a program to create, append, and remove lists in Python.
17. Write a program to demonstrate working with tuples in Python.
18. Demonstrate the following operators in Python with suitable examples:

    i) Arithmetic Operators ii) Relational Operators iii) Assignment Operator
    iv) Logical Operators v) Bitwise Operators vi) Ternary Operator
    vii) Membership Operators viii) Identity Operators.
19. Write a Python class to reverse a string word by word.

20. Demonstrate the following conditional statements in Python with suitable examples:

    i) if statement ii) if-else statement   iii) if-elif-else statement

21. Demonstrate the following iterative statements in Python with suitable examples:

    i) while loop   ii) for loop

22. Python program to perform read and write operations on a file.
23. Python program to count frequency of characters in a given file.
24. Write a Python program to return multiple values at a time using a return statement.
25. Demonstrate lambda functions in Python with suitable example programs.

SEMESTER – II

Course Code: MJ-MDS24-201
Course Title: Data Mining

Course Outcomes

On completion of the course, students will be able to:
CO1 – To understand the process of data understanding and cleaning.
CO2 – To familiarize with various data mining functionalities and how they can be applied to various real-world problems.
CO3 – To familiarize with various machine learning algorithms used in data mining.

Unit 1: Data Understanding and Classification Techniques

Data understanding and data cleaning, concept of supervised and unsupervised learning, problem of classification, classification techniques: k-nearest neighbor, decision tree, Naïve Bayesian, classification based on logistic regression.

Unit 2: Model Evaluation and Selection

Metrics for evaluating classifier performance, holdout method and random subsampling, cross-validation, bootstrap, model selection using statistical tests of significance, comparing classifiers based on cost–benefit and ROC curves. Techniques to improve classification accuracy: introduction to ensemble methods, bagging, boosting and AdaBoost, random forests.

Unit 3: Artificial Neural Networks (ANN) and Support Vector Machines (SVM)

Artificial Neural Network (ANN): Introduction to ANN, types of activation function, single layer network, multilayer feedforward network model, training methods, ANN and regression models. Support Vector Machine (SVM): introduction to support vector machine, loss functions, soft margin, optimization hyperplane, support vector classification, support vector regression, linear programming support vector machine for classification and regression.

Unit 4: Unsupervised Learning and Market Basket Analysis

Clustering: K-MEANS, k-medoids, CLARA. Market Basket Analysis: association rules and prediction, Apriori algorithm, data attributes, examples on association rule mining.

References:

1. Berson, S., & Smith, S. J. (1997). Data Warehousing, Data Mining, and OLAP. McGraw-Hill.
2. Breiman, J., Friedman, R. A., Olshen, C. J., & Stone, C. J. (1984). Classification and Regression Trees. Wadsworth and Brooks/Cole.
3. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd Edition.
4. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.

5. Ripley, B. D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press.
6. Mehrika, K., Mohan, C., & Ranka (1997). Elements of Artificial Neural Networks. Penram International.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.

Course Code: MJ-MDS24-202

Course Title: Programming in R

Course Outcomes

On completion of the course, students will be able to:

CO1 – To analyze and configure R software for statistical programming environment and describe generic programming language concepts implemented in a high-level statistical language.

CO2 – To demonstrate the programs in the R environment to create custom analytical models to meet the dynamic business needs.

CO3 – To evaluate and verify the analysis findings by using various packages in R Programming.

CO4 – To visualize and customize the various graphical packages for creating various types of graphs, plots, and charts.

CO5 – To review advanced data science concepts using predictive analytics fundamentals.

CO6 – To appraise and verify the analysis findings by conducting various statistical tests.

Unit 1: Introduction to R

What is R, Why R, Advantages of R over other programming languages. R Data Types: Vectors, Lists, Matrices, Arrays, Factors, Data Frame. R Variables: Variable assignment, Data types of variable, Finding variables , Deleting variables. R Operators: Arithmetic operators, Relational operators, Logical operators, Assignment operators, Miscellaneous operators. R Decision Making: if statement, if–else statement, if–else if statement, switch statement. R Loops: repeat loop, while loop, for loop; Loop control statements: break statement, next statement.

Unit 2: R Functions, Strings, Vectors, Lists, Matrices, and Arrays

R Functions: Function definition, built-in functions (mean(), paste(), sum(), min(), max(), seq()), user-defined functions, calling a function without arguments, calling a function with argument values. R Strings: Manipulating text data using substr(), strsplit(), paste(), grep(), toupper(), tolower(). R Vectors: Sequence vector, rep function, vector access, vector names, vector math, vector recycling, vector element sorting. R Lists: Creating a list, list tags and values, add/delete elements, size of list, merging lists, converting list to vector. R Matrices: Accessing elements of a matrix. R Arrays: Naming columns and rows, accessing array elements, manipulating array elements.

Unit 3: Data Frames and Handling Data in R

Data Frames: Create data frame, data frame access, understanding data in data frames using dim(), nrow(), ncol(), str(), summary(), names(), head(), tail(), edit(). Expanding data frames: add column, add row, joining columns and rows in a data frame using rbind() and cbind(), merging data frames using merge(), melting and casting data using melt() and cast(). Loading and handling data in R: getting and setting the working directory using getwd(), setwd(), dir().

Unit 4: CSV Files, Graphs, Statistical Functions, and RStudio

R CSV Files: Input as a CSV file, reading a CSV file, analyzing the CSV file using summary(), min(), max(), range(), mean(), median(), apply(). Graphs: Pie charts with titles and colors, slice percentages, chart legend, 3D pie charts, histograms, density plots, bar charts with labels, title, and colors. Built-in functions: lm, t.test, prop.test, wilcox.test, ks.test, var.test, chisq.test, aov. RStudio: R command prompt, R script file, comments. Handling Packages in R: Installing a package, commands like installed.packages(), packageDescription(), help(), find.package(), library(). Input and Output: Entering data from keyboard, printing fewer or more digits, handling special values: NA, Inf, -Inf.

References:

1. Uma Maheshwari, B., & Sujatha, R. (Year). Introduction to Data Science.
2. Vishwas R. Pawgi. Statistical Computing Using R Software.

Course Code: MJ-MDS24-203

Course Title: Time Series Analysis and Forecasting

Course Outcomes

On completion of the course, students will be able to:

CO1 – To understand forecasting techniques and familiarize with modern statistical methods for analyzing time series data.

CO2 – To develop problem-solving ability for addressing social issues and industrial problems.

Unit 1: Exploratory Analysis and Time Series Models

Exploratory analysis of time series: graphical display, classical decomposition model, components and various decompositions of time series models, numerical description of time series: stationarity, auto-covariance and autocorrelation functions, data transformations, methods of estimation, trend, seasonal and exponential smoothing techniques, moving average, exponential smoothing, Holt's and Winter's methods, exponential smoothing techniques for series with trend and seasonality, basic evaluation of exponential smoothing. Stationarity models: time series data, trend, seasonality, cycles and residuals, stationary, white noise processes, autoregressive (AR), moving average (MA), autoregressive and moving average (ARMA) and autoregressive integrated moving average (ARIMA) processes, choice of AR and MA periods. Non-stationarity models: tests for non-stationarity: random walk, random walk with drift, trend stationary, general unit root tests: Dickey-Fuller test, augmented Dickey-Fuller test. ARIMA models: basic formulation of the ARIMA model and their statistical properties, autocorrelation function (ACF), partial autocorrelation function (PACF) and their standard errors.

Unit 2: Forecasting and Model Evaluation

Forecasting: nature of forecasting, forecasting methods, qualitative and quantitative methods, steps involved in stochastic model building, forecasting model evaluation. Model selection techniques: AIC, BIC, and AICC – forecasting model monitoring. Transfer function and intervention analysis, transfer function models, transfer function, noise models, cross-correlation function, model specification, forecasting with transfer function and noise models, intervention analysis.

References:

1. Chatfield, C. (2001). *Time Series Forecasting*, Chapman & Hall, London.
2. Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd Ed., John Wiley.
3. Hamilton, N. Y. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
4. Kendall, Sir Maurice, & Ord, J. K. (1990). *Time Series* (Third Edition), Edward Arnold.

Course Code: GE-MDS24-204

Course Title: Exploratory Multivariate Data Analysis

Course Outcomes

On completion of the course, students will be able to:

CO1 – Understand and visualize multivariate data; apply PCA, factor analysis, and canonical correlation.

CO2 – Apply clustering methods and analyze properties of multivariate normal distributions.

CO3 – Perform multivariate regression, ANOVA/ANCOVA, and hypothesis testing using T^2 and D^2.

CO4 – Build and evaluate logistic regression and discriminant analysis models using classification metrics.

Unit 1: Multivariate Data and Dimension Reduction Techniques

Multivariate data and their diagrammatic representation, exploratory multivariate data analysis, sample mean vector, sample dispersion matrix, sample correlation matrix, graphical representation, mean, variance, covariance, correlation of linear transformations, six-step approach to multivariate model building. Introduction to multivariate regression models, principal component analysis (by using covariance and correlation method, standardized method), factor analysis (models, rotation types), canonical correlation with real-life examples.

Unit 2: Cluster Analysis and Multivariate Normal Distribution

Cluster analysis (hierarchical and non-hierarchical, agglomerative, single, complete, average, Wald's linkage, K-means clustering method, qualitative method clustering), multivariate normal distribution, singular and non-singular normal distribution, mean and variance of multivariate normal distribution, random sampling from multivariate normal distributions, independence of variables, M.G.F., characteristic function, moments.

Unit 3: Multivariate Linear Models and Hypothesis Testing

Multivariate linear model and analysis of variance and covariance, maximum likelihood estimation of parameters, tests of linear hypothesis, distribution of partial and multiple correlation coefficients and regression coefficients, multivariate linear regression, multivariate analysis of variance of one- and two-way classification data (only LR test), multivariate analysis of covariance, Hotelling's T^2 and Mahalanobis D^2 applications in testing and confidence set construction.

Unit 4: Logistic Regression and Discriminant Analysis

Logistic regression model and analysis: regression with a binary dependent variable, representation of the binary dependent variable, estimating the logistic regression model, assessing the goodness of fit of the estimation model, testing for significance of the coefficients, interpreting the coefficients, criteria for evaluation of logistic regression model – KS, Gini, AUC, Precision, Recall, F1 score, etc. Discriminant model and analysis: two-group discriminant analysis, three-group discriminant analysis, the decision process of discriminant analysis (objective, research design, assumptions, estimation of the model, assessing overall fit of a model, interpretation of results, validation of results).

References:

1. Anderson, T. W. (1984). *Introduction to Multivariate Analysis*, John Wiley.
2. Fang, K., Kotz, S., & Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall.
3. Härdle, W. K., & Simar, L. (2012). *Applied Multivariate Statistical Analysis*, Springer, New York.
4. Härdle, W. K., & Hlávka, Z. (2007). *Multivariate Statistics: Exercises and Solutions*.
5. Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). *Continuous Multivariate Distributions, Volume 1: Models and Applications*, John Wiley & Sons.
6. Kshirsagar, A. M. (1983). *Multivariate Analysis*, Marcel Dekker.
7. Morrison, D. F. (1990). *Multivariate Statistical Methods*, McGraw Hill Co.

Course Code: GE-MDS24-204

Course Title: Database Management System

Course Outcomes

On completion of the course, students will be able to:

CO1 – Be familiar with the fundamentals of database concepts and database management systems.

CO2 – Utilize conceptual modelling techniques, such as the ER model and relational model, to model the data requirements for an application.

CO3 – Write SQL commands to create tables, insert, update, delete, and query data.

Unit 1: Introduction to Databases and ER Modeling

Database system applications, database system vs file system, view of data, data abstraction, instances and schemas, data models, the ER model, relational model, database languages (DDL, DML), database access for application programs, database users and administrators, transaction management, database system structure, storage manager, the query processor.

Database design and ER diagrams – beyond ER design: entities, attributes, and entity sets; relationships and relationship sets; additional features of the ER model; conceptual design with the ER model.

Unit 2: Relational Model and Relational Algebra

Introduction to the relational model, integrity constraints over relations, enforcing integrity constraints, querying relational data, logical database design, introduction to views, destroying/altering tables and views.

Relational algebra – selection and projection, set operations, renaming, joins, division, examples of algebra queries. Relational calculus – tuple relational calculus, domain relational calculus.

Unit 3: SQL and Advanced Querying

The form of a basic SQL query, examples of basic SQL queries, introduction to nested queries, correlated nested queries, set-comparison operators, aggregate operators, NULL values, comparisons using NULL values, logical connectives (AND, OR, NOT) and their impact on SQL constructs, outer joins, disallowing NULL values.

Complex integrity constraints in SQL, triggers, views, procedures, and active databases.

Unit 4: Normalization and Transaction Management

Schema refinement, problems caused by redundancy, decompositions, problems related to decomposition, functional dependencies, reasoning about FDs, first, second, and third normal forms, BCNF, lossless join decomposition, dependency-preserving decomposition, schema refinement in database design, multivalued dependencies, fourth normal form.

Overview of transaction management: ACID properties, transactions and schedules, concurrent execution of transactions, lock-based concurrency control, performance of locking, transaction support in SQL.

Concurrency control: 2PL, serializability and recoverability, introduction to lock management, lock conversions, dealing with deadlocks, concurrency control without locking.

References:

1. Korth, Silberschatz. *Database System Concepts.*
2. Ivan Bayross. *SQL–PL/SQL*, BPB Publications.
3. Osborne. *Structured Query Language.*
4. O’Reilly. *Learning MySQL.*
5. Ivan Bayross. *SQL, PL/SQL: The Programming Language*, BPB.
6. Osborne. *Structured Query Language.*
7. Scott Ullman. *SQL.*
8. Dr. P. S. Deshpande. *SQL & PL/SQL Black Book for Oracle.*

Course Code: GE-MDS24-204

Course Title: Bayesian Inference

Course Outcomes

On completion of the course, students will be able to:

CO1 – Understand Bayesian inference, prior/posterior distributions, and decision-making using loss functions.

CO2 – Apply various types of priors and assess robustness and sensitivity in Bayesian analysis.

CO3 – Perform Bayesian model selection and analyze properties of posterior distributions.

CO4 – Implement Bayesian computational methods using EM, MCMC, Metropolis-Hastings, and Gibbs sampling.

Unit 1: Fundamentals of Bayesian Inference

Subjective and frequentist probability, Bayesian inference setup, prior and posterior distributions, loss functions, principles of minimum expected posterior loss, quadratic and other loss functions, advantages of being Bayesian, improper priors, common problems of Bayesian inference, point estimation, HPD confidence intervals, predictions of future observations, Bayesian testing.

Unit 2: Priors and Robustness in Bayesian Analysis

Bayesian analysis with subjective priors, classes of priors, conjugate class of priors, Jeffrey’s prior, probability matching prior, robustness, and sensitivity.

Unit 3: Bayesian Model Selection and Posterior Properties

Bayesian model selection, BIC, Bayes factors, limit of posterior distributions, consistency and asymptotic normality of posterior distributions.

Unit 4: Bayesian Computation

Bayesian computing, EM algorithm, MCMC, Metropolis–Hastings algorithms, Gibbs sampling, convergence diagnostics.

References

1. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press/Taylor & Francis. ISBN: 9781439840955.
2. Albert, J. (2009). *Bayesian Computation with R* (2nd ed.). Springer. ISBN: 0387922970.
3. Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods.* Springer, New York.
4. Gelman, A. et al. (2013). *Bayesian Data Analysis.* CRC Press.

Course Code: MJ-MDSP24-206

Course Title: Practical – II

1. K-Nearest Neighbor
2. Evaluation Measures Using Confusion Matrix
3. Artificial Neural Network and Support Vector Regression
4. Association Rule Mining
5. Data Input in R
6. Diagrammatic Representation in R
7. Graphical Representation in R
8. Measures of Central Tendency in R
9. Measures of Dispersion
10. Correlation in R
11. Regression in R
12. Testing of Hypothesis
13. Control and Loops in R
14. Estimation of Components of Time Series
15. Modelling of Time Series
16. Forecasting Time Series
17. Implement Data Definition Language (DDL) Commands
18. Implement Constraints
19. Implement Data Manipulation Language (DML) Commands
20. Implement Data Control Language (DCL) Commands
21. Computations Done on Data
22. Nested Queries and Joins
23. Views
24. Functions, Procedures, and Triggers

SEMESTER – III

Course Code: MJ-MDS24-301

Course Title: Machine Learning

Course Outcomes:

On completion of the course, the student will be able to:

CO1 – Explain fundamental concepts, types, and models of machine learning.

CO2 – Apply supervised and unsupervised algorithms such as decision trees, regression, SVMs, neural networks, and clustering methods.

CO3 – Implement ensemble and probabilistic models like Random Forest, Adaboost, Gaussian Mixture Models, and the EM algorithm.

CO4 – Analyze reinforcement and evolutionary learning methods such as Q-learning and genetic algorithms.

UNIT I: Introduction to Machine Learning

Introduction to Machine Learning – Introduction, Components of Learning, Learning Models, Geometric Models, Probabilistic Models, Logic Models, Grouping and Grading, Designing a Learning System, Types of Learning – Supervised, Unsupervised, Reinforcement; Perspectives and Issues, Version Spaces, PAC Learning, VC Dimension.

UNIT II: Supervised and Unsupervised Learning Algorithms

Supervised and Unsupervised Learning – Decision Trees: ID3, Classification and Regression Trees; Regression: Linear Regression, Multiple Linear Regression, Logistic Regression; Neural Networks: Introduction, Perceptron, Multilayer Perceptron; Support Vector Machines: Linear, Non-Linear, Kernel Functions; K-Nearest Neighbors; Introduction to Clustering: K-Means, K-Mode Clustering.

UNIT III: Ensemble and Probabilistic Learning Methods

Ensemble and Probabilistic Learning – Model Combination Schemes, Voting, Error-Correcting Output Codes, Bagging: Random Forest Trees, Boosting: Adaboost, Stacking; Gaussian Mixture Models, Expectation-Maximization (EM) Algorithm, Information Criteria, Nearest Neighbour Methods, Nearest Neighbour Smoothing, KD-Tree, Distance Measures.

UNIT IV: Reinforcement Learning and Evolutionary Algorithms

Reinforcement Learning and Evaluating Hypotheses – Learning Task, Q-Learning, Non-deterministic Rewards and Actions, Temporal-Difference Learning, Relationship to Dynamic Programming, Active Reinforcement Learning, Generalization in Reinforcement Learning; Basics of Sampling Theory, Error Estimation, Bias and Variance;

Genetic Algorithms – Motivation, Representation of Hypotheses, Genetic Operators, Fitness Function and Selection, Hypothesis Space Search, Genetic Programming, Models of Evolution and Learning: Lamarckian Evolution, Baldwin Effect, Parallelizing Genetic Algorithms.

References:

1. Tom Mitchell (1997), *Machine Learning*, McGraw Hill, 3rd Edition.
2. Stephen Marsland (2015), *Machine Learning: An Algorithmic Perspective*, 2nd Edition.
3. Christopher Bishop (2006), *Pattern Recognition and Machine Learning*, Springer.
4. Kevin P. Murphy (2012), *Machine Learning: A Probabilistic Perspective*, MIT Press.
5. Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009), *The Elements of Statistical Learning*, Springer.
6. Ethem Alpaydin (2014), *Introduction to Machine Learning*, 3rd Edition, MIT Press / Prentice Hall of India.

Course Code: MJ-MDS24-302

Course Title: Regression and Predictive Modeling

Course Outcomes:
At the end of the course, students will be able to:
CO1 – Deep comprehension of the linear and nonlinear regression models.
CO2 – Demonstrate understanding of model selection and regression modeling approaches.
CO3 – Examine the connections between dependent and independent variables.
CO4 – Estimate the parameters and fit a model.
CO5 – Investigate possible diagnostics in regression modeling and analysis.
CO6 – Validate the model using hypothesis testing and confidence interval approach.

Unit 1: Simple Linear Regression Analysis
Simple linear regression model, Ordinary Least Squares (OLS) method, generalized and weighted least squares, validating simple regression model using t-test and F-test, developing confidence intervals.

Unit 2: Multiple Linear Regression Analysis
Concept of multiple regression model, OLS method, generalized and weighted least squares, assessing the fit of the regression line, inferences from multiple regression analysis, problem of overfitting, comparing two regression models, prediction with multiple regression equation.

Unit 3: Model Adequacy Checking and Transformation Techniques
Residual analysis, PRESS statistics, detection and treatment of outliers, lack of fit of the regression model, tests of lack of fit, problems of autocorrelation and heteroscedasticity. Variance stabilizing transformations, transformations to linearize the model, Box-Cox methods, transformations on the regressors. Multicollinearity: sources, effects, diagnostics (correlation matrix, VIF, eigen system analysis of X'X), and methods of dealing with multicollinearity.

Unit 4: Generalized Linear Model and Logistic Regression
Generalized linear models: Link functions (normal, binomial, Poisson, exponential, gamma). Logistic regression: logit transform, maximum likelihood estimation, hypothesis tests (Wald test, LR test, score test), test for overall regression.

References:

1. Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). John Wiley.
2. Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley.
3. McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman

and Hall.

4. Montgomery, D. C., Peck, E. A., & Geoffrey, G. (2003). *Introduction to Linear Regression Analysis*. Wiley Eastern.
5. Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied Linear Statistical Models*.
6. Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling*. Marcel Dekker.

Course Code: MJ-MDS24-303

Course Title: Big Data

Course Outcomes:
At the end of the course, the student will be able to:
CO1 – Demonstrate knowledge of Big Data, data analytics, challenges, and their solutions.
CO2 – Analyze Hadoop framework and its ecosystem.
CO3 – Compare and work in NoSQL environments, including MongoDB.
CO4 – Apply Big Data using MapReduce programming in both Hadoop and Spark frameworks.

Unit 1: Foundations and Technologies of Big Data
Introduction to Big Data: Introduction to data, types of digital data (structured, semi-structured, unstructured), characteristics of data, sources of data.
Big Data Evolution: Definition, characteristics, need, and challenges of Big Data.
Big Data Analytics and Business Intelligence: Overview and importance.
Hadoop Ecosystem: Requirement of Hadoop framework, design principles, Hadoop architecture and components, comparison with traditional systems (SQL and RDBMS), Hadoop 1 vs. Hadoop 2.
MapReduce and YARN: Introduction to MapReduce, processing data with Hadoop using MapReduce, introduction to YARN, YARN architecture, managing resources and applications with Hadoop YARN.

Unit 2: Big Data Storage, Processing, and Analytics Tools
NoSQL Databases: Introduction, features and types, advantages and disadvantages, applications of NoSQL.
MongoDB: Introduction and features, data types, MongoDB Query Language, CRUD operations, arrays and functions (count, sort, limit, skip, aggregate, MapReduce), cursors and indexes, Mongo Import and Mongo Export.
Data Warehousing and Analytics with Hive: Introduction and architecture, data types and file formats, Hive Query Language (HQL) statements, partitions and bucketing, views and subqueries, joins, aggregations, group by, having, RCFile implementation, user-defined functions (UDFs), serialization and deserialization.
Modern Big Data Framework – Apache Spark: Introduction, how Spark works, programming with RDDs (creating RDDs, transformations, and actions), introduction to DataFrames and their use in Spark.

References:

1. White, T. (2015). *Hadoop: The Definitive Guide* (4th ed.). O’Reilly.
2. Guller, M. (2015). *Big Data Analytics with Spark*. Apress.
3. Miner, D., & Shook, A. (2012). *MapReduce Design Pattern*. O’Reilly.

Course Code: GE-MDS24-304

Course Title: Natural Language Processing

Course Outcomes:

On completion of the course, the student will be able to:

CO1 – Understand some of the problems and solutions of NLP and their relation to linguistics and statistics.

CO2 – Understand linguistic phenomena and learn to model them with formal grammars.

CO3 – Carry out proper experimental methodology for training and evaluating empirical NLP systems.

CO4 – Manipulate probabilities and construct statistical models over strings and trees.

CO5 – Estimate parameters using supervised and unsupervised training methods.

CO6 – Design, implement, and analyze NLP algorithms and different language modeling techniques.

Unit 1: Introduction to Natural Language Processing

Introduction, applications and use cases of NLP, components of NLP, steps in NLP.
Finding the structure of words: words and their components, lexemes, morphemes, morphology, problems in morphological processing, typology, morphological typology.
Natural Language Processing with Python NLTK package (text preprocessing tasks): word tokenization, sentence tokenization, filtering stop words, stemming, tagging parts of speech, lemmatization, chunking, chinking, named entity recognition, term frequency and inverse document frequency (TF-IDF).

Unit 2: Syntax Analysis

Parsing natural language, treebanks: a data-driven approach to syntax.
Representation of syntactic structure: syntax analysis using dependency graphs, syntax analysis using phrase structure trees.
Parsing algorithms: shift-reduce parsing, hypergraphs and chart parsing (CYK parsing).
Models for ambiguity resolution in parsing: probabilistic context-free grammar, generative models, discriminative models for parsing.

Unit 3: Language Modeling

Introduction, N-gram models, language model evaluation, parameter estimation, language model adaptation.
Types of language models and language-specific modeling problems.

Unit 4: Semantic Parsing

Introduction, semantic interpretation, system paradigms, word sense systems, software.
Predicate-argument structure, meaning representation systems, software.
Discourse processing: cohesion, reference resolution, discourse cohesion and structure.

References:

1. Indurkhya, N., & Damerau, F. J. (2010). *Handbook of Natural Language Processing* (2nd ed.). CRC Press/Taylor & Francis.
2. Jurafsky, D., & Martin, J. H. (2013). *Speech and Language Processing* (2nd ed.). Pearson Education India.
3. Manning, C., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
4. Bird, S., & Loper, E. (2016). *Natural Language Processing with Python* (2nd ed.). O'Reilly Media.

Course Code: GE-MDS24-304

Course Title: Cyber Security

Course Outcomes:

On completion of the course, the student will be able to:
CO1 – Analyze and evaluate the cyber security needs of an organization.
CO2 – Understand cyber security regulations and roles of international law.
CO3 – Design and develop a security architecture for an organization.
CO4 – Understand fundamental concepts of data privacy attacks.

Unit 1: Introduction to Cyber Security

Basic cyber security concepts, layers of security, vulnerability, threat, harmful acts, internet governance – challenges and constraints, computer criminals, CIA triad, assets and threats, motives of attackers, types of attacks: active, passive, software attacks, hardware attacks, cyber threats: cyber warfare, cyber crime, cyber terrorism, cyber espionage, comprehensive cyber security policy.

Unit 2: Cyberspace, Law, and Cyber Forensics

Cyber security regulations, roles of international law, the Indian cyberspace, National Cyber Security Policy, historical background of cyber forensics, digital forensics science, the need for computer forensics, cyber forensics and digital evidence, forensic analysis of email, digital forensics lifecycle, forensic investigation, challenges in computer forensics.

Unit 3: Cybercrime in Mobile, Wireless Devices, and Organizational Implications

Proliferation of mobile and wireless devices, trends in mobility, credit card frauds in mobile and wireless computing era, security challenges posed by mobile devices, registry settings, authentication service security, attacks on mobile/cell phones, organizational security policies and measures in mobile computing era, laptops, cost of cybercrimes and IPR issues, web threats for organizations, security and privacy implications of social media marketing, social computing and associated organizational challenges.

Unit 4: Privacy Issues and Case Studies

Basic data privacy concepts, data privacy attacks, data linking and profiling, privacy policies and their specifications, privacy policy languages, privacy in different domains: medical, financial, etc., cybercrime examples and mini-cases: official website of Maharashtra Government hacked, Indian banks lose millions of rupees, Parliament attack, Pune city police bust Nigerian racket, e-mail spoofing instances, mini-cases: online gambling in India, intellectual property crime, financial frauds in cyber domain.

References:

1. Preston Gralla, *How Personal and Internet Security Work,* Que Publications,
2. Alfred Basta, Wolf Halton, *Computer Security Concepts, Issues and Implementation*, Cengage Learning,
3. Joseph Pelton, Indu B. Singh, *Digital Defence: A Cybersecurity Primer,*
4. William Stallings, *Cryptography and Network Security: Principles and Practice*,
5. John R. Vacca, *Computer and Information Security Handbook.*

Course Code: GE-MDS24-304

Course Title: Digital Image Processing

Course Outcomes:

On completion of the course, students will be able to:

CO1 – Understand the fundamentals of digital images, visual perception, sampling, quantization, and color models.

CO2 – Apply spatial and frequency domain techniques for image enhancement and filtering.

CO3 – Perform image segmentation using edge detection, thresholding, region-based, and morphological methods.

CO4 – Analyze image compression techniques and implement image recognition and restoration methods.

Unit 1: Digital Image Fundamentals

Steps in digital image processing, components, elements of visual perception, image sensing and acquisition, image sampling and quantization, relationships between pixels, color image fundamentals: RGB and HSI models, two-dimensional mathematical preliminaries, 2D transforms.

Unit 2: Image Enhancement

Spatial Domain: Gray level transformations, histogram processing, basics of spatial filtering, smoothing and sharpening spatial filtering.

Frequency Domain: Introduction to Fourier Transform, smoothing and sharpening frequency domain filters, ideal, Butterworth, and Gaussian filters, homomorphic filtering, color image enhancement.

Unit 3: Image Segmentation

Edge detection, edge linking via Hough transform, thresholding, region-based segmentation: region growing, region splitting and merging, morphological processing: erosion and dilation, segmentation by morphological watersheds: basic concepts, dam construction, watershed segmentation algorithm.

Unit 4: Image Compression, Recognition, and Restoration

Image Compression: Need for data compression, Huffman coding, Run Length Encoding, Shift codes, Arithmetic coding, JPEG standard, MPEG.

Image Recognition: Boundary representation, boundary description, Fourier descriptors, regional descriptors, topological features, texture, patterns and pattern classes, recognition based on matching.

Image Restoration: Degradation model, properties, noise models, mean filters, order statistics, adaptive filters, band reject filters, band pass filters, notch filters, optimum notch filtering, inverse filtering, Wiener filtering.

References:

1. Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, Pearson, 4th Edition.
2. Anil K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.
3. Milan Sonka, Vaclav Hlavac, Roger Boyle, *Image Processing, Analysis, and Machine Vision*, Cengage Learning, 3rd Edition.
4. Wilhelm Burger, Mark J. Burge, *Digital Image Processing: An Algorithmic Introduction using Java*, Springer, 2009.
5. Bernd Jahne, *Digital Image Processing*, Springer, 6th Edition.

Course Code: MJ-MDSP24-306

Course Title: Practical-III

1. Implement a simple Linear Regression model to predict house prices using a sample dataset.
2. Perform Logistic Regression on a dataset (e.g., Titanic survival or Iris dataset) and evaluate performance.
3. Build a Decision Tree classifier (ID3 or CART) on the Iris dataset and visualize the tree.
4. Implement K-Nearest Neighbors (KNN) classifier for handwritten digit recognition (MNIST dataset).
5. Implement K-Means clustering on a dataset and visualize results.
6. Implement Support Vector Machine (SVM) classifier on Iris dataset with different kernel functions.
7. Implement Random Forest classifier on Titanic dataset and evaluate feature importance.
8. Implement Adaboost classifier on the Breast Cancer dataset and compare with Decision Tree.
9. Implement Principal Component Analysis (PCA) for dimensionality reduction.
10. Implement Hierarchical Clustering and draw a dendrogram.
11. Implement Naïve Bayes Classifier for text classification.
12. Implement Logistic Regression with Regularization (L1 vs L2).
13. Simple Linear Regression.
14. Model Adequacy Checking.
15. Multicollinearity.
16. Generalized Linear Model.
17. Data Types in MongoDB.
18. CRUD Operations & Query Language in MongoDB.
19. Arrays and Functions in MongoDB.
20. Cursors, Indexes in MongoDB.

SEMESTER – IV

Course Code: MJ-MDS24-401
Course Title: Deep Learning

Course Outcomes:
CO1 – Recognize the characteristics of deep learning models useful to solve real-world problems.
CO2 – Identify and apply appropriate deep learning algorithms for analyzing data for a variety of problems.
CO3 – Implement different deep learning algorithms.

Unit I: Introduction to Deep Learning
Definition and Applications, Neural Networks, Machine Learning vs Deep Learning, Deep Learning Libraries: TensorFlow, Keras, PyTorch, Types of Deep Learning: Supervised Learning, Unsupervised Learning, Reinforcement Learning (and their comparison), Datasets: Numerical Data, Categorical Data, Data Quality and Remediation, Data Preprocessing: Dimensionality Reduction, Feature Transformation, Feature Subset Selection.

Unit II: Neural Networks
Basics, Types, and Intuitions, Neurons, Kernels, Biases, Weights, Initialization, Gradient Descent and Heuristics, Training Methods: Holdout, K-Fold Cross-Validation, Bootstrap Sampling, Lazy vs Eager Learner, Evaluation: Regression, Classification, Clustering, Perceptrons, Derivatives, Computation Graph, Vectorization, Broadcasting, Propagation: Forward and Backward, Parameters vs Hyper parameters.

Unit III: Deep Feed forward Network
Feed forward Networks and Gradient-based Learning, Hidden Units and Architecture Design, Computational Graphs and Back propagation, Regularization, Parameter Penalties, Data Augmentation, Multitask Learning, Bagging, Dropout, Adversarial Training, Optimization Techniques.

Unit IV: Convolution Networks and Sequence Modelling
Convolution Networks: Convolution Operation, Pooling, Basic Convolution Function, Convolution Algorithm, Unsupervised Features and Neuroscientific Insights for Convolution Networks, Sequence Modelling: Recurrent Neural Networks (RNNs), Bidirectional RNNs, Encoder-Decoder Sequence-to-Sequence Architectures, Deep Recurrent Networks, Recursive Neural Networks, Echo State Networks.

References:
1. Goodfellow, L., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, 2016,
2. Patterson, J., & Gibson, A., *Deep Learning: A Practitioner's Approach*, O'Reilly, 1st Edition, 2017,
3. Haykin, S., *Neural Networks and Machine Learning*, Prentice Hall Pearson, 3rd Edition, 2009,
4. Geron, A., *Hands-on Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, 2017.

Course Code: MJ-MDS24-402
Course Title: Data Visualization

Course Outcomes:
CO1 – Understand the fundamentals, principles, and cognitive foundations of data visualization.
CO2 – Apply Python and R libraries to create static, statistical, and interactive visualizations.
CO3 – Design professional dashboards using Power BI and apply business intelligence techniques.
CO4 – Develop interactive dashboards using advanced Power BI features such as DAX, bookmarks, and drill-through.

Unit 1: Fundamentals of Data Visualization

Importance of visualization in analytics; principles of perception & cognition; types of data – categorical, numerical, time-series, geospatial; core charts – line, bar, scatter, histogram, pie; advanced charts – heatmap, treemap, bubble chart; choosing the right visualization; color theory and design basics; introduction to storytelling with data.

Unit 2: Visualization in Python and R

Python Libraries: Matplotlib – custom charts, subplots, annotation ; Seaborn – statistical plots, pairplots, distribution charts; Plotly – interactive dashboards; Bokeh – interactive graphics; GeoPandas + Folium – geospatial visualizations.

R Libraries: ggplot2 – grammar of graphics, facets, themes; Plotly-R – interactive charts; Shiny – basic dashboards.

Unit 3: Business Intelligence – Power BI Basics

Getting started with Power BI Desktop – interface, importing data, connecting multiple sources; Power Query – data cleaning, transformation, merging, and appending datasets; data modeling – star schema, relationships, cardinality, role-playing dimensions; core visuals – KPI cards, tables, matrices, slicers, charts, maps; hands-on – simple dashboards with multiple visuals and filters.

Unit 4: Business Intelligence – Power BI Advanced

Advanced DAX – calculated columns, measures, ranking, filtering; dashboard/report design – layout, themes, bookmarks, drill-through, tooltips; publishing dashboards – Power BI Service, sharing, scheduled refresh, workspace management; hands-on – building 2–3 full interactive dashboards with real datasets.

References

1. Microsoft Power BI Documentation & Guided Learning
2. Cole Nussbaumer Knaflic – *Storytelling with Data*
3. Tamara Munzner – *Visualization Analysis & Design*

4. Wes McKinney – *Python for Data Analysis*
5. Hadley Wickham – *ggplot2: Elegant Graphics for Data Analysis*

Course Code: MJ-MDS24-403
Course Title: Blockchain Technology & Applications

Course Outcomes:
CO1 – Understand the fundamentals, architecture, and types of blockchain,
CO2 – Explore cryptography, consensus mechanisms, and security in blockchain,
CO3 – Develop smart contracts and decentralized applications,
CO4 – Analyze real-world blockchain applications across industries

Unit I: Introduction to Blockchain

Blockchain Fundamentals: Definition, history, and evolution, Features: decentralization, immutability, transparency, Types of Blockchain: Public, private, consortium, Blockchain Architecture: Blocks, chains, nodes, ledger structure, Applications of Blockchain: Cryptocurrency, supply chain, healthcare, finance.

Unit II: Cryptography and Security in Blockchain

Cryptography Basics: Hash functions, SHA-256, public/private keys, Digital Signatures: Signing transactions, verification, Security Principles: Consensus mechanisms (PoW, PoS), Attack vectors (51% attack, double spending)

Unit III: Smart Contracts and Decentralized Applications (DApps)

Smart Contracts: Definition, characteristics, use cases, Platforms: Ethereum, Hyperledger Fabric, Developing DApps: Solidity basics, contract deployment, Interaction with blockchain using Web3.js, Testing and Security: Vulnerabilities, auditing contracts.

Unit IV: Blockchain Applications and Future Trends

Enterprise Blockchain Solutions: Hyperledger, Corda, Integration with business systems, Blockchain in Finance and Supply Chain: Payments, remittances, asset tracking, Emerging Trends: NFTs, DeFi, Web3, Scalability and interoperability solutions.

References:

1. Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S., *Bitcoin and Cryptocurrency Technologies*, Princeton University Press, 2016
2. Crosby, M., Pattanayak, P., Verma, S., & Kalyanaraman, V., *Blockchain Technology: Beyond Bitcoin*, Applied Innovation Review, 2016
3. Mougayar, W., *The Business Blockchain*, Wiley, 2016
4. Antonopoulos, A. M., *Mastering Bitcoin*, O’Reilly, 2017
5. Wood, G., *Ethereum: A Secure Decentralised Generalised Transaction Ledger*, 2014
6. Alharby, M., & van Moorsel, A., *Blockchain-Based Smart Contracts: A Systematic Mapping Study*, 2017

Course Code: GE-MDS24-403

Course Title: Cloud Computing

Course Outcomes:

CO1 – To distinguish the different models and computing paradigms.

CO2 – To explain the levels of virtualization and resource virtualization.

CO3 – To analyze the reasons for migrating into cloud.

CO4 – To effectively use cloud services in terms of infrastructure and operating platforms.

CO5 – To apply the services in the cloud for real-world scenarios.

Unit I: Basics of Cloud Computing

Overview, Applications, Intranets and the Cloud, Your Organization and Cloud Computing, Benefits, Limitations, Security Concerns,Software as a Service (SaaS): Multitenant Nature of SaaS Solutions, Understanding SOA,Platform as a Service (PaaS): IT Evolution Leading to the Cloud, Benefits and Disadvantages of PaaS Solutions,Infrastructure as a Service (IaaS): Understanding IaaS, Load Balancing, System and Storage Redundancy, Cloud-Based NAS, Advantages, Server Types.

Unit II: Data Storage, Security, and Virtualization in Cloud

Cloud File Systems: GFS, HDFS, BigTable, HBase, Dynamo,Cloud Data Stores: Data Store, Simple DB, Overview of Cloud Storage, Cloud Storage Providers,Securing the Cloud: Security Advantages, Business Continuity, Disaster Recovery, Threats,Virtualization: Introduction, Types (Application, Network, Desktop, Storage, Server, Data),Cloud Deployment Models: Public, Private, Community, Hybrid,Role of Cloud Computing in Data Science, Advantages in Machine Learning.

Unit III: Cloud Service Providers and Applications

Amazon Web Services: EC2, S3, SQS, EBS, ELB, Simple DB, RDS, Virtual Amazon Cloud,

Google Cloud: AppEngine, Google Storage,Microsoft Azure, Rackspace Cloud,Cloud Applications: Business and Consumer Applications (CRM, ERP, Productivity, Social Networking, Media, Multiplayer Gaming, E-Commerce), Cloud for eGovernance, Scientific Applications (Healthcare, Biology, Geoscience).

Unit IV: Cloud Computing for Data Science

Machine Learning in the Cloud: Benefits and Limitations,Types of Cloud-Based Machine Learning Services: AIaaS, GPUaaS, Managed ML Platforms,Cloud

Machine Learning Platforms: AWS SageMaker, Azure Machine Learning Studio, Google Cloud AutoML.

References:

1. Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi, *Mastering Cloud Computing*, McGraw Hill, 2013.
2. Thomas Erl, Zaigham Mahmood, Ricardo Puttini, *Cloud Computing: Concepts, Technology & Architecture*, Prentice Hall, 2013.
3. George Reese, *Cloud Application Architectures*, O'Reilly Media, 2009.
4. Barrie Sosinsky, *Cloud Computing Bible*, Wiley, 2011.

Course Code: GE-MDS24-403C

Course Title: Business Intelligence and Data Analytics

Course Outcomes:

CO1 – Explain the role of business intelligence and analytics in a dynamic business environment.
CO2 – Apply modern tools for statistical modeling and visualization.
CO3 – Demonstrate descriptive analytics for business intelligence and data warehousing.
CO4 – Implement algorithms for data mining techniques and processes.
CO5 – Develop scripts for text & web mining and social network analysis.

Unit I: Overview of Business Intelligence, Analytics, Data Science, and AI

Changing business environments and evolving needs for decision support, Decision-making processes, Computerized decision support frameworks, Evolution from decision support to analytics/data science, Framework for business intelligence and analytics overview, Artificial Intelligence concepts, drivers, major technologies, and business applications, Conversational AI—Chatbots.

Unit II: Descriptive Analytics I – Nature of Data, Big Data, and Statistical Modeling

Nature of data in analytics, Taxonomy of data, Data preprocessing, Definition of Big Data, Fundamentals and technologies of Big Data analytics, Big Data and stream analytics, Statistical modeling for business analytics, Regression modeling for inferential statistics.

Unit III: Descriptive Analytics II – Business Intelligence, Data Warehousing, and Visualization

Business intelligence and data warehousing, Data warehousing process and architectures, Data management and warehouse development, Warehouse administration, Security issues and future trends, Business reporting, Data visualization, Charts and graphs, Visual analytics, Information dashboards.

Unit IV: Predictive and Prescriptive Analytics, Text, Web, and Social Media Analytics

Data mining concepts, applications, process, and methods, Optimization and simulation, Model-based decision-making, Decision modeling with spreadsheets, Text analytics and text mining overview, Natural Language Processing (NLP), Text mining applications and process, Sentiment analysis and topic modeling, Web mining overview, Search engines, Web usage mining (Web analytics), Social analytics.

References:

1. Steve Williams, *Business Intelligence Strategy and Big Data Analytics - A General Management Perspective*, Morgan Kaufmann (Elsevier), 2016.
2. Vincent Charles, Pratibha Garg, Neha Gupta, Mohini Agarwal, *Data Analytics and Business Intelligence - Computational Frameworks, Practices, and Applications*, CRC Press, 2023.
3. Ira J. Haimowitz, *Data Analytics for Business - Lessons for Sales, Marketing, and Strategy*, Routledge (Taylor & Francis), 2023.

Course Code: MJ-MDSP24-405

Course Title: Practical I

1. Implementation of Feed Forward Neural Network (FFNN) for Binary Classification
2. End-to-End Data Preprocessing Pipeline for Machine Learning and Deep Learning
3. Implementation of Gradient Descent Algorithm from Scratch
4. Implementation of Multi-Layer Perceptron (MLP) from Scratch
5. Hyperparameter Tuning in Deep Neural Networks
6. Cross-Validation and Bootstrap Evaluation for Model Performance Assessment
7. Convolutional Neural Network (CNN) for Image Classification
8. Feature Extraction and Visualization using Convolutional Neural Network (CNN)
9. Sequence Modeling using Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM)
10. Data Visualization Using Matplotlib
    (Create line, bar, and scatter plots,Use subplots,Add labels, legends, and annotations).
11. Statistical Visualization Using Seaborn
    (Create distribution plots, Generate pair plots, Draw box and violin plots).
12. Interactive Visualization Using Plotly (Python).
    (Create interactive scatter and bar charts,Use hover, zoom, and filters).
13. Data Visualization in R Using ggplot2
    (Create bar, line, and scatter plots,Apply facets and themes).
14. Data Cleaning and Transformation Using Power Query
    (Remove duplicates and null values,Change data types,Rename and split columns,Merge and append datasets).
15. Creating KPI Cards and Tabular Visuals in Power BI
    (Create KPI cards,Build tables,Create matrix visuals,Format tabular data).
16. Creating Charts and Slicers in Power BI
    (Add slicers for filtering,Create bar charts,Create line charts,Create pie charts).
17. Case Study –Business Dashboard
    (Load a business dataset (sales/finance) ,Clean and model the data ,Create KPIs and charts ,Build an interactive dashboard with slicers).
18. Creating Calculated Columns Using DAX
    (Create calculated columns ,Use basic DAX functions ,Perform row-level calculations).
19. Ranking and Filtering Using DAX
    (Rank data using DAX ,Apply Top N filtering,Create conditional measures).
20. Advanced Dashboard and Report Design
    (Design report layout,Apply themes and formatting,Use bookmarks,Create drill-through pages).