M.Sc. CS (Big Data Analytics)

(2 Years Programme)

Syllabus

Department of Data Science and Analytics

School of Mathematics, Statistics and Computational Sciences

Central University of Rajasthan

NH-8 Jaipur- Ajmer Highway, Bandarsindri

Kishangarh -305817

District-Ajmer, Rajasthan

Website: www.curaj.ac.in

M.Sc. CS (Big Data Analytics)

M.Sc. CS (Big Data Analytics)

Program Outcomes (POs)

PO1	Understand theoretical aspects of data science, machine learning and statistical methods used in Big Data analysis, deploy real time machines on cloud.
PO2	Explain and adhere to the processes involved in conducting research that results in a quality dissertation and contributes to society.
PO3	Apply the understanding of big data concepts through practical implementations.
PO4	Evaluate the knowledge of Big Data Analytics in the frontier areas of research and allied areas.
PO5	Prepare trained data analyst and manpower needed for academics, research and industry.
PO6	Prepare trained individuals to tackle emerging research and analytics issues in computing domain in general and also with a focused approach on big data domain.
PO7	Facilitate practical knowledge and problems solving skills to approach newer technological problems in the big data domain, and computing domain in general.

Program Specific Outcomes (PSOs)

PSO1	At the end of the program, graduates will be able to get insights into various fields of Big Data Analytics with a deep understanding of theoretical aspects of security and related analysis.
PSO2	Graduates should also get a broader understanding of various aspects of big data, statistics behind the data science, prediction standards, practical applicability, and their limitations.
PSO3	During the course, students should enhance their inquisitiveness to ever-evolving domain of Big Data Analytics and apply their knowledge to solve problems.
PSO4	With a focused one semester project-based internship, student will be able to work on real life projects and get exposure of industries.
PSO5	Students should also be able to solve prediction-based issues with understanding of Big Data and time-series attributes with a relevance to standards.

ANNEXURE-III

Scheme & Syllabus
M.Sc. CS (Big Data Analytics) (2 Years Programme)

Semester- I

S. No.	Subject Code	Course Title	Type of Course	L	T		P	Credit
1	MBD 401	STATISTICAL METHODS	CC	3	0		2	4
2	MBD 402	PROBABILITY DISTRIBUTIONS	CC	3	0		2	4
3	MBD 403	LINEAR ALGEBRA AND MATRIX THEORY	CC	3	0		0	3
4	MBD 404	COMPUTING FOR DATA SCIENCES	CC	2	0		4	4
5	MBD 405	DATABASE MANAGEMENT SYSTEM	CC	2	0		4	4
6	MBD 406	PROFESSIONAL COMMUNICATION	CC	2	0		0	2
7	MBD 407	PYTHON AND JAVA	CC	1	0		4	3
		Total Credits						**24**

Semester- II

S. No.	Subject Code	Course Title	Type of Course	L	T	P	Credit
1	MBD 408	FOUNDATIONS OF DATA SCIENCE	CC	2	0	4	4
2	MBD 409	ADVANCED STATISTICAL METHODS	CC	3	0	2	4
3	MBD 410	MACHINE LEARNING	CC	2	0	4	4
4	MBD 411	VALUE THINKING	CC	2	0	4	4
5	MBD 412	COMBINATORIAL OPTIMIZATION	CC	3	0	2	4
6	MBD 413	INTRODUCTION TO ECONOMETRICS AND FINANCE	CC	3	1	0	4
		Total Credits					**24**

Semester- III

S. No.	Subject Code	Course Title	Type of Course	L	T	P	Credit
1	MBD 501	**MODELING** IN OPERATIONS MANAGEMENT	CC	2	0	4	4
2	MBD 502	ENABLING TECHNOLOGIES FOR DATA SCIENCE	CC	2	0	4	4
3	MBD 503	DATA MINING	CC	3	0	2	4
4	MBD 504	CLOUD COMPUTING	CC	3	0	2	4
5		ELECTIVE-I	DE	3	0	2	4
6		ELECTIVE- II	DE	3	0	2	4
		Total Credits					**24**

Semester- IV

S. No.	Subject Code	Course Title	Type of Course	L	T	P	Credit
1	MBD 581	INTERNSHIP BASED PROJECT	AECC	-	-	-	24
		Total Credits					**24**

CC: Compulsory Course, DE: Department Elective, AECC: Ability Enhancement Compulsory Course

Semester-I

Programme: M.Sc. Big Data Analytics, Semester-I (2022-2023)

<table>
<tr><td colspan="3">Course: STATISTCAL METHODS
Code: MBD 401</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of statistics, Data Collection and Visualization</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Describe the datasets including both categorical and quantitative variables.
• Apply laws of probability to concrete problems.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Data Collection: Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with date collection systems with examples from business, cleaning and treatment of missing data.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Data Visualization: Principles of data visualization, and different method of presenting data in business analytics.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Basic Statistics: Frequency table, histogram, measures of location, measures of spread, skewness, kurtosis, percentiles, box plot, correlation and simple linear regression, partial correlation, probability distribution as a statistics model, fitting probability distributions, empirical distributions, checking goodness of fit through plots and tests.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Contingency Tables: Two way contingency tables, measures of association, testing for dependence.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. Statistics: David Freedman, Pobert Pisani & Roger Purves, W. W. Norten & Co. 4th Edition 2007.
Reference Books:
1. The visual display of Quantitative information: Edward Tufte, Graphics Press, 2001
2. Best Practices in Data Cleaning: Jason W. Osborne, Sage Publications 2012.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-I (2022-2023)

Course: PROBABILITY DISTRIBUTIONS **Code: MBD 402**		
TEACHING SCHEME:	**EXAMINATION SCHEME:**	**CREDITS ALLOTED:**
Theory: 3 **Tutorial: 0** **Practical: 2**	**Internal Assessment: 40 Marks** CIA-I: 20 Marks (Unit I & II) CIA-II: 20 Marks (Unit III) **End Semester Examination: 60 Marks** ESE: Unit I to IV	**4**
Course Pre-requisites: UG Level Calculus		
Course Objectives: To develop the knowledge of probability and its distributions		
Course Outcomes: After completing the course, the student shall be able to: • Explain the concept of random events. • Express the features of discrete and continuous random variables. • Formulate the distribution functions. • Define the discrete/continuous distributions and solve the problems about these distributions.		
Course Content:		
Unit-I	**Discrete Distributions:** Binomial, Poisson, multinomial, hypergeometric, negative binomial, uniform. The (a,b,0) class of distributions. Moments, quantiles, cdf, survival function and other properties.	**(15 Hours)**
Unit-II	**Continuous Distributions:** Uniform, Normal, Exponential, gamma, Weibull, Pareto, lognormal, Laplace, Cauchy, Logistic distributions; properties and applications. Functions of random variables and their distributions using Jacobian of transformation and other tools. Bivariate normal and bivariate exponential distributions.	**(15 Hours)**
Unit-III	Concept of a sampling distribution. Sampling distributions of t, $\chi 2$ and F (central and noncentral), their properties and applications. Multivariate normal distribution. Distribution of linear function of normal random variables. Characteristic function of the multivariate normal distribution. Distribution of quadratic forms. Cochran's theorem. Independence of quadratic forms.	**(15 Hours)**
Unit-IV	Compound, truncated and mixture distributions. Convolutions of two distributions. Order statistics: their distributions and properties. Joint, marginal and conditional distribution of order statistics. The distribution of range and median. Extreme values and their asymptotic distribution (statement only) with applications.	**(15 Hours)**

Text Book:

1. Rohatgi V.K & A.K. MD. Ehsanes Saleh: An Introduction to Probability Theory and Mathematical Statistics, 2nd. John Wiley and Sons, 2001.

Reference Books:

1. Rao, C.R. (2002). Linear Statistical Inference and its Applications, (2nd Ed.Wiley).
2. Reiss R D, Thomas M. Statistical Analysis of Extreme Values with application to Insurance, Finance, Hydrology and other Fields (Second Edition) BirkhäuserVerlag
3. Chandra T.K., Chatterjee D., A first course in Probability (2001) Narosa Publication.

Programme: M.Sc. Big Data Analytics, Semester-I (2022-2023)

<table>
<tr><td colspan="3">Course: LINEAR ALGEBRA AND MATRIX THEORY
Code: MBD 403</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>3</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of matrices, Linear transformations, Eignen values and vectors</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Solve systems of linear equations using multiple methods, including Gaussian elimination and matrix inversion.
• Carry out matrix operations, including inverses and determinants.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Fields, System of Linear Equations. Matrices, Elementary Row Operations, Row Reduced Matrices, Invertible Matrices, Vector spaces, Subspaces, Linear Combinations, Linear span, Linear dependence and Linear independence of vectors, Basis and Dimension, Ordered Basis, Finite dimensional vector spaces, Sum and Direct sum of subspaces</td><td>(12 Hours)</td></tr>
<tr><td>Unit-II</td><td>Linear transformations and their representation as matrices, Kernel and Image of a linear transformation, Rank and Nullity Theorem, Change of Basis</td><td>(11 Hours)</td></tr>
<tr><td>Unit-III</td><td>Eigen values and Eigen vectors of a linear transformation (matrices), Characteristic polynomial and minimal polynomial, Diagonalization of linear operators, invariant subspaces, Jordon form</td><td>(11 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Inner product spaces, Cauchy-Schwarz-inequality, Orthogonal vectors, Orthonormal sets and bases, Positive Definite and Positive Semi-definite Matrices, Linear functional on an Inner Product Space and Adjoint of a Linear operator, Self-adjoint operators, Normal Operators, Spectral theorem.</td><td>(11 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. Linear Algebra, Kenneth Hoffman and Ray Kunze, Prentice Hall.
Reference Books:
1. Linear Algebra, Friedberg, Pearson Education.
2. Linear Algebra and Applications, Gilbert Strang, Academic Press.
3. Linear Algebra Done Right, Axler, Sheldon, Springer International Publishing.
4. Finite Dimensional Vector Spaces, P.R. Halmos, Springer-Verlag, New York.
5. Linear Algebra: A Geometric Approach, S. Kumaresan</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-I (2022-2023)

<table>
<tr><td colspan="3">Course: COMPUTING FOR DATA SCIENCES
Code: MBD 404</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 4</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of R and Python packages, Concepts of computation and Computing methodologies.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Develop the ability to build and assess data-based models.
• Execute statistical analysis with professional statistical software.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Computer Package: Usage of R & Python with illustration.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Concepts of Computation: Algorithms, Convergence, Complexity with illustrations..</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Sorting & searching algorithms, numerical methods e.g. Newton Raphson, Steepest ascent</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Computing Methodologies: Monte-Carlo simulations of random numbers and various statistical methods, memory handling strategies for big data.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. Software for Data Analysis – Programming with R: John M. Chambers, Springer
Reference Books:
1. Elementary Numerical Analysis – An Algorithmic Approach: Samuel Conte and Carl de Boor (McGraw-Hill Education)
2. 2. Introduction to Algorithms: Cormen, Leiserson, Rivest and Stein, The MT Press (Third Edition)</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-I (2022-2023)

<table>
<tr><td colspan="3">Course: DATABASE MANAGEMENT SYSTEM
Code: MBD 405</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 4</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of Data Base and SQL queries</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
Draw Entity-Relationship diagrams to represent simple database application scenarios
Write SQL queries for a given context in relational database.
Discuss normalization techniques with simple examples.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Basic Concepts: Need, purpose and goal of DBMS, Three tier architecture, ER Diagram, data models- Relational, Network, Hierarchical. Database Design: Conceptual data base design, concept of physical and logical databases, data abstraction and data independence, data aggregation</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Relational data base: Relations, Relational Algebra, Theory of Normalization, Functional Dependency, Primitive and Composite data types</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Application Development using SQL: DDL and DML, Host Language interface, embedded SQL programming, Stored procedures and triggers and views, Constraints assertions. NoSQL Databases</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Internal of RDBMS: Physical data organization in sequential, indexed random and hashed files. Inverted and multilist structures, B trees, B+ trees, Query Optimization, Join Algorithm, Statistics and Cost Base optimization. Parallel and distributed data base. Transaction Processing, concurrency control, and recovery management. Transaction model properties and state serializability. Lock base protocols, two phase locking.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
Database system concepts : Abraham Silberschartz, Henry F. Korth and S. Surarshan, McGraw Hill, 2011.
Reference Books:
Almasri and S.B. Navathe: Fundamentals of Database Systems, C.J. Date: Data Base Design, Addison Wesley
Hansen and Hansen : DBM and Design, PHI</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-I (2022-2023)

<table>
<tr><td colspan="3">Course: PROFESSIONAL COMMUNICATION
Code: MBD 406</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>2</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of communication theory and processes.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Understand and apply communication theory.
• Critically think about communication processes and messages.
• Write effectively for a variety of contexts and audiences.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Grammar and Vocabulary: Tenses, subject–verb agreement. Sentence Analysis: Simple, Compound and Complex sentences. Phrases: Adjective, Adverb and Noun Phrase, Clauses: Adjective, Adverb and Noun Phrase. Voice, Narration, Gerund, Participle.</td><td>(7 Hours)</td></tr>
<tr><td>Unit-II</td><td>Oral Communication: Listening Skill – Active listening, Barriers to active listening, Speaking Skill- Stress patterns in English, Questioning skills, Barriers in Speaking, Reading Skill- Skimming, Scanning, Intensive reading, linking devices in a text, Different versions of a story/ incident</td><td>(8 Hours)</td></tr>
<tr><td>Unit-III</td><td>Written communication: Writing process, paragraph organization, writing styles, Types of Writing - Technical vs. creative; Types of technical writing, Scientific Writing: Writing a Scientific Report</td><td>(7 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Soft Skills: Body Language – Gesture, posture, facial expression. Group Discussion – Giving up of PREP, REP Technique. Presentation Skills: (i) How to make power point presentation (ii) Body language during presentation (iii) Resume writing: Cover letter, career objective, Resume writing (tailor made). Interview Skills: Stress Management, Answering skills.</td><td>(8 Hours)</td></tr>
<tr><td colspan="3">Text Books:
1. Advanced English Usage: Quirk & Greenbaum; Pearson Education.
2. Developing Communication Skills: Banerjee Meera & Mohan Krishna; Macmillan Publications, 1990.
3. Business Communication: Chaturvedi, P.D.; Pearson Publications.
Reference Books:
1. Business Communication; Mathew, M.J.; RBSA Publications, 2005.
2. Communication of Business; Taylor, Shirley; Pearson Publications.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-I (2022-2023)

<table>
<tr><td colspan="3">Course: PYTHON AND JAVA
Code: MBD 407</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 4</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>3</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of programming.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• To acquire programming skills in core Python and Java.
• Acquire Object Oriented Skills in Python and Java.
• Develop the skill of designing Graphical user Interfaces in Python and Java.
• Develop the ability to write database applications in Python and Java.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Introduction to Python: The basic elements of python, Branching Programs, Control Structures, Strings and Input, Iteration, Functions, Scoping and Abstraction, Functions and scoping, Specifications, Recursion, Global variables, Modules ,Files , System Functions and Parameters</td><td>(10 Hours)</td></tr>
<tr><td>Unit-II</td><td>Python Data Structures: Structured Types, Mutability and Higher-Order Functions, Strings, Tuples, Lists and Dictionaries, Lists and Mutability, Functions as Objects, Testing, Debugging, Exceptions and Assertions, Types of testing – Black-box and Glass-box, Debugging, Handling Exceptions, Assertions. Simple Algorithms and Data structures, Search Algorithms, Sorting Algorithms, Hash Tables. Object Oriented Python: Classes and Object-Oriented Programming, Abstract Data Types and Classes, Inheritance, Encapsulation and Information Hiding,
Regular Expressions – REs and Python, Plotting using PyLab</td><td>(12 Hours)</td></tr>
<tr><td>Unit-III</td><td>Introduction to Java: Overview and characteristics of Java, Java program compilation and execution process, Organization of JVM, JVM as interpreter and emulator, sandbox model. Data Types, primitive variables, arrays, operators, Control statements, standard input-output and main method. Object Oriented concepts: Concept of encapsulation and abstraction, Designing Classes, objects, instance variables and methods, Class modifiers, Inheritance, Interfaces, Abstract classes, Polymorphism: overloading and overriding, Composition.</td><td>(10 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Constructors: use of this and super, Java Stack and Heap, Garbage collection. Static methods and variables, Wrapper classes, Autoboxing, Standard classes: Math and String
Exception Handling & applications: exception types, nested try-catch, throw, throws and finally statements. Multithread Programming: thread creation, synchronization and priorities. Input-output and file operations: Java.io, Object serialization and deserialization. Java Collections API: Arraylist, Set, list, Map, Hashtable, Comparator and comparable Database connectivity, Java Packages, creating a jar file</td><td>(13 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. John V Guttag. “Introduction to Computation and Programming Using Python”, Prentice Hall of India.
2. Programming with Java (2019). E Balagurusamy, McGraw-Hill; 6th edition.
Reference Books:</td></tr>
</table>

1. Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser, “Data Structures and Algorithms in Pyhon”, Wiley.
2. Java A Beginner's Guide (2020). Herbert Schildt, McGraw Hill; 8th Edition.

Semester-II

Programme: M.Sc. Big Data Analytics, Semester-II (2022-2023)

<table>
<tr><td colspan="3">Course: FOUNDATIONS OF DATA SCIENCE
Code: MBD 408</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 4</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of data collection, processing, analysis and visualization.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Cover the technical pipeline from data collection, to processing, analysis, and visualization.
• Collecting data, sampling, quality assessment, and repair.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>High Dimensional Space: Properties, Law of large number, Sphere and cube in high dimension, Generation points on the surface of sphere, Gaussians in high dimension, Random projection, Applications.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Random Graphs: Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.
Singular Value Decomposition (SVD): Best rank k approximation, Power method for computing the SVD, PCA.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Random Walks and Markov Chains: Properties of random walks, Stationary distributions, Random walks on undirected graphs with unit edge weights, Random walks in Euclidean space, Markov Chain Monte Carlo. Algorithm for Massive Data Problems, Frequency moments of data streams, Matrix algorithms using sampling</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>The General Models for Massive Data Problems: Topic Models - Non-Negative Matrix Factorization, Latent Dirichlet Allocation (LDA), Hidden Markov Models, Graphical Models and Belief Propagation, Bayesian Networks, Markov Random Fields.</td><td>(15 H ours)</td></tr>
<tr><td colspan="3">Text Book:
1. Foundation of Data Science: Avrim Blum, John Hopcroft, and Ravindran Kannan.
Reference Books:
1. Introduction to Data Science: Practical Approach with R and Python (2021). B. Uma Maheswari Wiley India Pvt Ltd., 1st Edition.
2. Data Science and Machine Learning using Python (2022). Reema Thareja. McGraw Hill Education (India) Private Ltd., 1st Edition.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-II (2022-2023)

<table>
<tr><td colspan="3">Course: ADVANCED STATISTICAL METHODS
Code: MBD 409</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of sample and population, hypothesis testing.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Explain the difference between sample and population.
• Explain the difference between a statistics and a parameter.
• Draw the hypothesis testing decision matrix and explain the contents.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Estimation: Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Test of Hypotheses: Two types of errors, test statistic, parametric tests for equality of means & variances.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Gauss Markov Model, least square estimators, Analysis of variance.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Regression: Multiple linear regression, forward, backward & stepwise regression, Logistic regression</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. Mathematical Statistics, Vol. I (2000): P. J. Bickel and K.A. Docksum, Prentice Hall, 2nd Edition.
Reference Book:
1. Introduction to Linear Regression Analysis (2007): Douglas C. Montgomery. Wiley–Blackwell; 4th edition.
2. Linear Regression: A Tutorial Introduction to the Mathematics of Regression Analysis (2022): James V Stone.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-II (2022-2023)

<table>
<tr><td colspan="3">Course: MACHINE LEARNING
Code: MBD 410</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 4</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of machine learning algorithms, data analytics solution.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
Appreciate the importance of visualization in the data analytics solution.
Apply structured thinking to unstructured problems.
Understand a very broad collection of machine learning algorithms and problems.
Learn algorithmic topics of machine learning and mathematically deep enough to introduce the required theory.
Develop an appreciation for what is involved in learning from data.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Basics: Introduction to Machine Learning, Different Forms of Learning.
Regression Analysis: Linear Regression, Ridge Regression, Lasso, Bayesian Regression, Regression with Basis Functions. Classification Methods: Instance-Based Classification, Linear Discriminant Analysis, Logistic Regression, Large Margin Classification, Kernel Methods, Support Vector Machines, Multi-class Classification, Classification and Regression Trees.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Neural Networks: Multi-layer Networks, Back-propagation, Multi-class Discrimination, Training Procedures, Localized Network Structure, Deep Learning.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Graphical Models: Hidden Markov Models, Bayesian Networks, Markov Random Fields, Conditional Random Fields.
Ensemble Methods: Boosting - Adaboost, Gradient Boosting, Bagging - Simple Methods, Random Forest.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Clustering: Partitional Clustering - K-Means, K-Medoids, Hierarchical Clustering - Agglomerative, Divisive, Distance Measures, Spectral Clustering.
Dimensionality Reduction: Principal Component Analysis, Independent Component Analysis, Multidimensional Scaling, and Manifold Learning.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
Pattern Recognition and Machine Learning (2016): Christopher Bishop, Springer, 1st Edition.
Reference Books:
Machine Learning (1997): Tom Mitchell, McGraw-Hill International Editions, 1st Edition.
Machine Learning: An Artificial Intelligence Approach (1995): Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell, Morgan Kaufmann, 1st Edition.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-II (2022-2023)

Course: VALUE THINKING **Code: MBD 411**		
TEACHING SCHEME:	**EXAMINATION SCHEME:**	**CREDITS ALLOTED:**
Theory: 2 **Tutorial: 0** **Practical: 4**	**Internal Assessment: 40 Marks** CIA-I: 20 Marks (Unit I & II) CIA-II: 20 Marks (Unit III) **End Semester Examination: 60 Marks** ESE: Unit I to IV	**4**

Course Pre-requisites: No

Course Objectives: To develop the knowledge of explaining an issue, analyzing context.

Course Outcomes: After completing the course, the student shall be able to:

- Explaining an issue or problem.
- Analyzing context.
- Describing your and others perspective.

Course Content:

This course involves watching few movies (list provided below) such as Twelve Angry Men, Roshman, and reading few books (list provided below) that deals mostly with argumentative logic, evidence, drawing inference from evidences. After watching each movie and reading each book, there will be general discussion amongst the students. Each student will prepare a term paper. Evaluation will be on the basis of this term paper and participation in group discussion.

Unit-I	Movie to watch- Twelve Angry Men, Discussion about the argumentative logic, evidence, drawing inference from evidences. Book to read- The Hound of the Baskervilles by Arthur Conan Doyle	**(15 Hours)**
Unit-II	Movie to watch- Roshoman by Kurosawa, Discussion about the argumentative logic, evidence, drawing inference from evidences. Book to read- Five Little Pigs by Agatha Christie	**(15 Hours)**
Unit-III	Movie to watch- Trial of Nuremberg, Discussion about the argumentative logic, evidence, drawing inference from evidences. Book to read- The Purloined Letter by Edger Allan Poe	**(15 Hours)**
Unit-IV	Movie to watch- Mahabharata by Peter Brook, Discussion about the argumentative logic, evidence, drawing inference from evidences. Book to read- The Case of the Substitute Face	**(15 Hours)**

Text and Reference Books:

1. The Hound of the Baskervilles by Arthur Conan Doyle
2. Five Little Pigs by Agatha Christie
3. The Purloined Letter by Edger Allan Poe
4. The Case of the Substitute Face

Programme: M.Sc. Big Data Analytics, Semester-II (2022-2023)

<table>
<tr><td colspan="3">Course: COMBINATORIAL OPTIMIZATION
Code: MBD 412</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of linear programming and designing efficient algortihms.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Familiarize with linear programming techniques for designing efficient algorithms.
• Design and analyze algorithms using the techniques.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Linear Optimization Problem: Linear Programming, Simplex method, Revised Simplex method, Duality, Dual Simplex, Interior Point Method, Transportation problem, Assignment Problem</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Non-linear Optimization Problem: General Non-Linear Unconstrained Optimization, convex function, concave function, local & global optimum, quadratic programming</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Discreet Optimization Problem: Local search, Greedy Algorithm, Dynamic Programming, Branch & Bound Algorithm, Network Flow Problem: Shortest Path Problem, Knapsack problem, Max-Flow and Min-cut problem</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Combinatorial Optimization Problems in Computer vision, social networks, cyber physical systems, Big Data analytics</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. Combinatorial Optimization -Theory and Algorithms, Bemhard Korte, Jens Vygen.
Reference Books:
1. Combinatorial Optimization, W.J. Cook, W.H. Cunningham, W.R. Pulleyblank, A. Schrijver.
2. Introduction to Combinatorial Analysis (2002): John Riordan, Dover Publications Inc., 1st Edition.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-II (2022-2023)

<table>
<tr><td colspan="3">Course: INTRODUCTION TO ECONOMETRICS AND FINANCE
Code: MBD 413</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 1
Practical: 0</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of basic statistics, econometric computer package, interpret linear-regression.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
Comfortable with basic statistics and probability.
To use a statistical/econometric computer package to estimate an econometric model and be able to report the results of their work in a non-technical and literate manner.
Able to estimate and interpret linear regression models and be able to distinguish between economic and statistical importance.
Able to critique reported regression results and interpret the results for someone who is not trained as an economist.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Introduction to Econometrics (using finance concepts): Assumptions of Classical Linear Regression Model, Ordinary Least Squares approaches, Autocorrelation, Heteroscedasticity, Multi collinearity, Dummy Variable approaches, and Distributed lag models.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>A brief Introduction to Time Series and Panel data models, Components of time series, Stationary and non-stationary time series, ARMA and ARIMA models, Static panel data models: fixed effects and random effects.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Basics of Finance: Time value of money, concept of present and future value analysis, stock and bond valuations, risk and return, Systematic and unsystematic risk, Diversification, cost of capital, capital structure,</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Dividend Discount Model, Portfolio Theory, Efficient Market Hypothesis (EMH), Capital asset pricing model (CAPM), Market Volatility, Options.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Books:
Richard A. Brealey, Stewart C. Myers, Franklin Allen, Pitabas Mohanty (2012). Principle of Corporate Finance, (10th ed.), Tata McGraw Hill Education Private Limited
Chris Brooks (2014). Introductory econometrics for Finance, Cambridge University Press.
Reference Books:
Pamela Peterson Drake, Frank J. Fabozzi, (2010). The Basics of Finance: An Introduction to Financial Markets, Business Finance and Portfolio Management, John Wiley & Sons, Inc.
Jeffery m Wooldridge (2015). Introductory Econometrics: A Modern Approach, Cengage learning
Damodar N. Gujarati. Basic Econometrics, Tata McGraw Hill Education Private Limited</td></tr>
</table>

Semester-III

Programme: M.Sc. Big Data Analytics, Semester-III (2022-2023)

<table>
<tr><td colspan="3">Course: MODELING IN OPERATIONS MANAGEMENT
Code: MBD 501</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 4</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of input-output framework, roles and responsibilities of operations managers.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Understand the input–process–output framework, the extensions of it, and apply them to a wide range of operations.
• Examine the types of transformation processes occurring within operations.
• Define the roles and responsibilities of operations managers and the challenges they face.
• Reflect on your own operations management responsibilities, if applicable.
• Understand the content of an operations strategy and the decisions involved.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td colspan="3">The course involves training the students to design, manage and improve a firm’s systems and processes. The objective is to develop skills to combines data, technology, and mathematical models to help managers make better decisions, identify new opportunities, and become more competitive</td></tr>
<tr><td>Unit-I</td><td>Classify various operations management problems, Identify the nature of the information needed to be able to address the problem, translate these problems into the appropriate statistical and/or mathematical framework and interpret the results of the models in a verbal manner of the case study: Venture Analysis, Banking analytics</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Classify various operations management problems, Identify the nature of the information needed to be able to address the problem, translate these problems into the appropriate statistical and/or mathematical framework and interpret the results of the models in a verbal manner of the case study: Marketing analytics</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Classify various operations management problems, Identify the nature of the information needed to be able to address the problem, translate these problems into the appropriate statistical and/or mathematical framework and interpret the results of the models in a verbal manner of the case study: Healthcare analytics, Retail analytics</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Classify various operations management problems, Identify the nature of the information needed to be able to address the problem, translate these problems into the appropriate statistical and/or mathematical framework and interpret the results of the models in a verbal manner of the case study: Supply chain analytics</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. S. Anil Kumar & N. Suresh (2009). Operations Management. New Age International (P) Ltd., Publishers.
Reference Books:
1. William J. Stevenson (2002), Operations Management Models: A Problem-Solving Approach, McGraw-Hill Higher Education.
2. Introduction to Supply Chain Management (2015): Handfield and Nichols, Prentice Hall India Learning Private Limited; First edition.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-III (2022-2023)

<table>
<tr><td colspan="3">Course: ENABLING TECHNOLOGIES FOR DATA SCIENCE
Code: MBD 502</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 2
Tutorial: 0
Practical: 4</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of big data and its analytics, Hadoop and NOSQL.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
Understand Big Data and its analytics in the real world.
Analyze the Big Data framework like Hadoop and NOSQL to efficiently store and process Big Data to generate analytics.
Design of Algorithms to solve Data Intensive Problems using Map Reduce Paradigm.
Design and Implementation of Big Data Analytics using pig and spark to solve data intensive problems and to generate analytics.
Implement Big Data Activities using Hive</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Big Data and Hadoop: Hadoop architecture, Hadoop Versioning and configuration, Single node & Multi-node Hadoop, Hadoop commands, Models in Hadoop, Hadoop daemon, Task instance, illustrations.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Map-Reduce: Framework, Developing Map-Reduce program, Life cycle method, Serialization, Running Map Reduce in local and pseudo-distributed mode, illustrations.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>HIVE: Installation, data types and commands, illustration.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>SQOOP: Installation, importing data, Exporting data, Running, illustrations.
PIG: Installation, Schema, Commands, illustrations.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
Hadoop in Action: Chuck Lam, 2010, ISBN: 9781935182191.
Reference Books:
Data-intensive Text Processing with Map Reduce: Jimmy Lin and Chris Dyer, Morgan & Claypool Publishers, 2010.
Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning (2019). Raj Kamal and Preeti Saxena, McGraw Hill Education, 1st Edition.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-III (2022-2023)

<table>
<tr><td colspan="3">Course: DATA MINING
Code: MBD 503</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of attributes of dataset, relationships between two or more attributes.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
Find out how different attributes of dataset are related to each other through patterns and data visualization techniques.
Find out relationships between two or more attributes of a dataset, and use this to predict outcomes or actions.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Data Mining: Introduction, Techniques, Issues and challenges, applications, Data preprocessing, Knowledge representation.
Association Rule Mining: Introduction, Methods to discover association rules, Association rules with item constraints</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Decision Trees: Introduction, Tree construction principle, Decision tree construction algorithm, pruning techniques, Integration of pruning and construction; Cluster analysis: Introduction, clustering paradigms, Similarity and distance, Density, Characteristics of clustering algorithms, Center based clustering techniques, Hierarchical clustering, Density based clustering, other clustering techniques, Scalable clustering algorithms, Cluster evaluation</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>ROC Curves: Introduction, ROC Space, Curves, Efficient generation of Curves, Area under ROC Curve, Averaging ROC curves, Applications</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Advanced techniques: Web mining - Introduction, Web content mining, Web structure mining, Web usage mining; Text mining- Unstructured text, Episode rule discovery from text, Text clustering; Temporal data mining – Temporal association rules, Sequence mining, Episode discovery, time series analysis; Spatial data mining – Spatial mining tasks, Spatial clustering, Spatial trends.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
Data Mining: Concepts and Techniques. Jiawei Han and Michelline Kamber. Elsevier (2011).
Reference Book:
Mastering Data Mining (2000). M. Berry and G. Linoff, John Wiley & Sons.
High-Dimensional Probability:An Introduction with Applications in Data Science (2018). Roman Vershynin Cambridge University Press; 1st edition.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-III (2022-2023)

<table>
<tr><td colspan="3">Course: CLOUD COMPUTING
Code: MBD 504</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of various service models, changing infrastructure landscape worldwide.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Gain knowledge of various service models.
• How cloud systems is changing the infrastructure landscape worldwide.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Overview of Computing Paradigm, Recent trends in Computing Grid Computing, Cluster Computing, Distributed Computing, Utility Computing, Cloud Computing Evolution of cloud computing Business driver for adopting cloud computing.
Introduction to Cloud Computing. Cloud Computing (NIST Model) Introduction to Cloud Computing, History of Cloud Computing, Cloud service providers. Properties, Characteristics & Disadvantages Pros and Cons of Cloud Computing.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Cloud Computing Architecture, Comparison with traditional computing architecture (client/server), Services provided at various levels, How Cloud Computing Works, Role of Networks in Cloud computing, protocols used, Role of Web services</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Service Models (XaaS), Infrastructure as a Service (IaaS), Platform as a Service(PaaS), Software as a Service(SaaS),Deployment Models ,Public cloud, Private cloud, Hybrid cloud Community cloud, Infrastructure as a Service(IaaS). Introduction to virtualization, Different approaches to virtualization, Hypervisors, Machine Image, Virtual Machine (VM).</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Examples, Amazon EC2, Renting, EC2 Compute Unit, Platform and Storage, pricing, customers Eucalyptus, Platform as a Service (PaaS),Introduction to PaaS, What is PaaS, Service Oriented Architecture (SOA).
Examples: Google App Engine, Microsoft Azure, SalesForce.com’s Force.com platform Software as a Service (PaaS): Introduction to SaaS, Web services, Web 2.0, Web OS, Case Study on SaaS Cloud Security. Case Study on Open Source & Commercial Clouds.</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. Erl Thomas, Puttini Ricardo, Mahmood Zaigham (2014). Cloud Computing: Concepts, Technology & Architecture (2014), Pearson Education India.
Reference Book:
1. Cloud Computing Bible, Barrie Sosinsky, Wiley-India, 2010.
2. Cloud Computing: Master The Concepts, Architecture And Applications With Real-World Examples And Case Studies (2019). Kamal Kant Hiran, BPB Publications; 1st edition.</td></tr>
</table>

LEVEL-5 (Electives)			
S.No.	**Course Code**	**Course Title**	**Credit**
1	**MBD 531**	**TIME SERIES AND FORECASTING**	**4**
2	**MBD 532**	**MULTIVARIATE STATISTICS**	**4**
3	**MBD 533**	**BIOINFORMATICS**	**4**
4	**MBD 534**	**SOFTWARE ENGINEERING**	**4**

Course: TIME SERIES AND FORECASTING
Code: MBD 531

TEACHING SCHEME:	**EXAMINATION SCHEME:**	**CREDITS ALLOTED:**
Theory: 3 **Tutorial: 0** **Practical: 2**	**Internal Assessment: 40 Marks** CIA-I: 20 Marks (Unit I & II) CIA-II: 20 Marks (Unit III) **End Semester Examination: 60 Marks** ESE: Unit I to IV	**4**

Course Pre-requisites: No

Course Objectives: To develop the knowledge of time series with different structures, seasonality, and cyclical irregularity.

Course Outcomes: After completing the course, the student shall be able to:

- Explain time series with different structures.
- Explain trend, seasonality, cyclical irregularity.
- Construct and evaluate time series models.

Course Content:

Unit-I	**Basics of Time series:** A model Building strategy, Time series and Stochastic process, stationarity, Auto correlation, meaning and definition – causes of auto correlation - consequence of autocorrelation – test for auto – correlation. Study of Time Series model and their properties using correlogram, ACF and PACF. Yule walker equations	**(15 Hours)**
Unit-II	**Time Series Models:** White noise Process, Random walk, MA, AR, ARMA and ARIMA models, Box- Jenkins's Methodology fitting of AR(1), AR(2), MA(1), MA(2) and ARIMA(1,1) process. Unit root hypothesis, Co-integration, Dicky Fuller test unit root test, augmented Dickey – Fuller test.	**(15 Hours)**
Unit-III	Non-linear time series models, ARCH and GARCH Process, order identification, estimation and diagnostic tests and forecasting. Study of ARCH (1) properties. GARCH (Conception only) process for modelling volatility.	**(15 Hours)**
Unit-IV	**Multivariate Liner Time series**: Introduction, Cross covariance and correlation matrices, testing of zero cross correlation and model representation. Basic idea of Stationary vector Autoregressive Time Series with order one: Model Structure, Granger Causality, stationarity condition, Estimation, Model checking.	**(15 Hours)**

Text Books:

1. Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis – Forecasting and Control, Holden – day, San Francisco.
2. Ruey S. Tsay (2014). Multivariate Time series Analysis: with R and Financial Application, Wiley & Sons.

Reference Books:

1. Chatfield, C. (2003) Analysis of Time Series, An Introduction, CRC Press.
2. Ruey S. Tsay (2005). Analysis of Financial Time Series, Second Ed. Wiley & Sons.
3. 5. Introduction to Statistical Time Series : W.A. Fuller

Programme: M.Sc. Big Data Analytics, Semester-III (2022-2023)

<table>
<tr><td colspan="3">Course: MULTIVARIATE STATISTICS
Code: MBD 532</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of finding patterns and correlations between several variables simultaneously, analyze complex datasets.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
• Find patterns and correlations between several variables simultaneously.
• Analyze complex datasets, allowing to gain deeper understanding of data and how it relates to real-world scenario.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Review of Multivariate Normal Distribution (MVND) and related distributional results. Random sampling from MVND, Unbiased and maximum likelihood estimators of parameters of MVND, their sampling distributions, independence. Correlation matrix and its MLE. Partial and multiple correlation coefficients, their maximum likelihood estimators (MLE), Wishart distribution and its properties (only statement).</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Hotelling's T2 and its applications. Hotelling's T2 statistic as a generalization of square of Student's statistic. Distance between two populations, Mahalnobis D2 statistic and its relation with Hotelling's T2 statistic.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Classification problem – two populations, two multivariate normal populations, several populations; Discriminant analysis - Fischer’s method, Logistic Regression Principle component analysis – Introduction, population principal components, summarizing sample variation by principal components, graphing principal components.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Canonical correlation – Introduction, canonical variates & correlations, interpreting canonical variables,
Factor Analysis – Introduction, Orthogonal Factor model, Methods of Estimation, Factor Rotation & Scores, and Perspective & Strategy for Factor Analysis
Cluster Analysis – Introduction, similarity measures, hierarchical & non-hierarchical clustering methods, multidimensional scaling, correspondence analysis</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:
1. Kshirsagar A. M.(1972) : Multivariate Analysis. Maral-Dekker.
Reference Books:
1. Johnosn, R.A. and Wichern. D.W (2002): Applied multivariate Analysis. 5th Ed. Prentice –Hall.
2. Anderson T. W. (1984): An introduction to Multivariate statistical Analysis 2nd Ed. John Wiely.
3. Morrison D.F. (1976): Multivariate Statistical Methods McGraw-Hill.</td></tr>
</table>

Programme: M.Sc. Big Data Analytics, Semester-III (2022-2023)

Course: BIOINFORMATICS **Code: MBD 533**		
TEACHING SCHEME:	**EXAMINATION SCHEME:**	**CREDITS ALLOTED:**
Theory: 3 **Tutorial: 0** **Practical: 2**	**Internal Assessment: 40 Marks** CIA-I: 20 Marks (Unit I & II) CIA-II: 20 Marks (Unit III) **End Semester Examination: 60 Marks** ESE: Unit I to IV	**4**

Course Pre-requisites: No

Course Objectives: To develop the knowledge of basic principles and concept of biology, computer science and mathematics, extract information from large databases and use this information in computer modelling.

Course Outcomes: After completing the course, the student shall be able to:

- Knowledge and awareness of the basic principles and concepts of biology, computer science and mathematics.
- Existing software effectively to extract information from large databases and to use this information in computer modeling.

Course Content:

The objective is to train the students to learn to identify, compile, analyze, and communicate complex biological and genetic data for initiatives in areas such as human genome analysis, disease research, and drug discovery and development. The course involves study of existing algorithms and application of data analytics to biological data.

Unit-I	Sequence Alignment problem & Algorithm, Pairwise & Multiple sequence Alignment, Advance Alignment Method.	**(15 Hours)**
Unit-II	Gibbs Sampling, Population Genomics, Genetic Mapping, Disease Mapping	**(15 Hours)**
Unit-III	Gene Recognition, Transcriptome & Evolution, Protein Structure, Protein Motifs.	**(15 Hours)**
Unit-IV	Hidden Markov Models, Lattice Model, Algorithms.	**(15 Hours)**

Text Book:

1. Arthur Lesk (2014). Introduction to Bioinformatics, Oxford University Press, 4th Edition.

Reference Book:

1. Introduction Computational Molecular Biology: C Setubal & J Meidanis, PWS Publishing Boston, 1997.
2. Bioinformatics: A Concept-Based Introduction (2010). Venkatarajan Subramanian Mathura & Pandjassarame Kangueane, Springer, 1st Edition.

Programme: M.Sc. Big Data Analytics, Semester-III (2022-2023)

<table>
<tr><td colspan="3">Course: SOFTWARE ENGINEERING
Code: MBD 534</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical: 2</td><td>Internal Assessment: 40 Marks
CIA-I: 20 Marks (Unit I & II)
CIA-II: 20 Marks (Unit III)
End Semester Examination: 60 Marks
ESE: Unit I to IV</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Pre-requisites: No</td></tr>
<tr><td colspan="3">Course Objectives: To develop the knowledge of various phase of software lifecycle, choosing process model.</td></tr>
<tr><td colspan="3">Course Outcomes: After completing the course, the student shall be able to:
Decompose the given project in various phases of a lifecycle.
Choose appropriate process model depending on the user requirements.
Perform various life cycle activities like Analysis, Design, Implementation, Testing and Maintenance.
Know various processes used in all the phases of the product.
Apply the knowledge, techniques, and skills in the development of a software product.</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Software Development Life Cycle: Software Process, Software Development Life Cycle Models , Software Requirement Engineering: Requirement Engineering Process
Function-oriented Design: Introduction to Structured Analysis, Data Flow Diagram, Process Specification, Entity Relationship (ER) Model, Structured Design Methodologies, Design Metrics</td><td>(15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Object Oriented Concepts & Principles: Key Concepts, Relationships: Is-A Relationship, HasA Relationship, Uses-A Relationship; Modelling Techniques: Booch OO Design Model, Rumbaugh’s Object Modelling Technique, Jacobson’s model, The Unified Approach to Modelling, Unified Modelling Language (UML). Object Oriented Analysis & Design: UseCase Modelling, Use-Case Realization, Class Classification Approaches: Noun Phrase Approach, CRC Card Approach, Use-case Driven Approach, Identification of Classes, Relationship, Attributes and Method. System Context and Architectural Design, Principles of Class Design, Types of Design Classes</td><td>(15 Hours)</td></tr>
<tr><td>Unit-III</td><td>UML 2.0 diagrams: Structure diagrams, Behavior diagrams
Software coding and Testing: Coding standards and guidelines, Code review techniques, Testing Fundamentals, Verification & Validation, Black Box Testing, White Box Testing, Unit Testing, Integration Testing, System Testing, Object Oriented System Testing.</td><td>(15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Emerging Trends: Architecture styles, Service Oriented Architecture (SOA), CORBA, COM/DCOM; Web Engineering: General Web Characteristics, Emergence of Web Engineering, Web Engineering Process, Web Design Principles, Web Metrics</td><td>(15 Hours)</td></tr>
<tr><td colspan="3">Text Book:</td></tr>
</table>

1. "Software Engineering: A Practitioner's Approach", Roger S. Pressman, McGraw Hill, 6/e, 2005.

Reference Books:

1. Rajib Mall. Fundamentals of Software Engineering, PHI, 3rd Edition, 2009.
2. Grady Booch, James Rambaugh, Ivar Jacobson, Addison-Wesley. Unified Modeling Language Users Guide, 2nd Edition.

Semester-IV

Programme: M.Sc. Big Data Analytics, Semester-IV (2022-2023)

Course: INTERNSHIP BASED PROJECT **Code: MBD 581**		
TEACHING SCHEME:	**EXAMINATION SCHEME:**	**CREDITS ALLOTED:**
Theory: - **Tutorial: -** **Practical: -**	**Project evaluation of total 200 Marks** **Internal Evaluation (by 3-member including internal supervisor) – 100 Marks** **External Evaluation – 100 Marks**	**24**
Course Pre-requisites: No		
Course Objectives: To develop the knowledge real time industry experience.		
Course Outcomes: After completing the course, the student shall be able to: • Integrate theory and practice. • Assess interest and abilities in their field of study. • Learn to appreciate works and its functions in the economy.		
Course Content: A real-life project must be undertaken at an undertaken at an industry for 20 weeks. Each student will have two supervisors: Once from academic institution and one from the industry. The project shall involve handling data extensively and use of methodologies learnt during the course work to derive meaningful inferences. A final project report has to be submitted and an "open" presentation has to be made.		