ST JOSEPH'S UNIVERSITY, BENGALURU

A Private – Public Partnership University Under RUSA 2.0 of MHRD(Government of India) established by the Karnataka Govt. Act No. 24 of 2021

SCHOOL OF INFORMATION TECHNOLOGY

DEPARTMENT OF ADVANCED COMPUTING

SYLLABUS FOR POSTGRADUATE PROGRAMME

SUMMARY OF CREDITS IN MSc (BIG DATA ANALYTICS)

Department of Advanced Computing (PG) Revision Year - 2026						
Semester 1	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Max marks for SE duration of examination**
Theory	BDA 1125	Statistical and Probabilistic Methods	45	03	03	50-2Hrs
Theory	BDA 1225	Linear Algebra & Linear Programming	45	03	03	50-2Hrs
Theory	BDA 1325	Computing for Data Science	45	03	03	50-2Hrs
Theory	BDA 1425	Database Management System	45	03	03	50-2Hrs
Theory	BDA 1525	Research Methodology	30	02	02	30-1.5Hrs
Practical	BD1P1	Statistical and Probabilistic Methods Lab	30	02	01	50-2Hrs
Practical	BD1P2	Linear Algebra & Linear Programming Lab	30	02	01	50-2Hrs
Practical	BD1P3	Computing for Data Science Lab	30	02	01	50-2Hrs
Practical	BD1P4	Database Management System Lab	30	02	01	50-2Hrs
Total Number of credits:			**18**			
Semester 2	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Max marks for SE duration of examination**
Theory	BDA 2125	Foundation of data Science	45	03	03	50-2Hrs

Theory	BDA 2225	Advanced Analytics	45	03	03	50-2Hrs
Theory	BDA 2325	Machine Learning	45	03	03	50-2Hrs
Theory	BDA 2425	Enabling Technologies for Data Science I	45	03	03	50-2Hrs
Theory	BDA 2525	Value thinking	45	03	03	50-2Hrs
Theory	BDA 2625	Data Mining	45	03	03	50-2Hrs
Theory	BDA 2725	Digital Information Processing	45	03	03	50-2Hrs
Practical	BD2P1	Foundation of data Science Lab	30	02	01	50-2Hrs
Practical	BD2P2	Advanced Analytics Lab	30	02	01	50-2Hrs
Practical	BD2P3	Data Mining and Machine Learning Lab	30	02	01	50-2Hrs
Practical	BD2P4	Enabling Technologies for Data Science I Lab	30	02	01	50-2Hrs
Practical	BD2P5	Digital Information Processing Lab	30	02	01	50-2Hrs
Practical	BD2P6	Research Oriented Project/ Paper	30	02	01	50-2Hrs
Total Number of credits:			**26**			
Semester 3	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Max marks for SE duration of examination**
Theory	BDA 3125	Modeling in Operations Management	45	03	03	50-2Hrs
Theory	BDA 3225	Enabling Technologies for Data Science II	45	03	03	50-2Hrs
Theory	BDA 3325	Deep Learning	45	03	03	50-2Hrs
Theory	BDA 3425	Data Analytics on Cloud	45	03	03	50-2Hrs
Theory	BDA 3525	Introduction to Econometrics and Finance	45	03	03	50-2Hrs

Practical	BD3P1	Modeling in Operation Management Lab	30	02	01	50-2Hrs
Practical	BD3P2	Enabling Technologies for Data Science II Lab	30	02	01	50-2Hrs
Practical	BD3P3	Deep Learning Lab	30	02	01	50-2Hrs
Practical	BD3P7	Research Oriented Project/ Paper	120	08	04	100-4Hrs
Total Number of credits:			**23**			
Semester 4	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Max marks for SE duration of examination**
Practical	BDA4IN25	Internship/ project	600		20	100
Practical	BD4P1	Dissertation/Research Paper	180	12	6	100
		IGNITORS/ OUTREACH			03	
Total Number of credits:			**29**			

CORE COURSES (CC)	
Course Title	Code Number
Statistical and Probabilistic Methods	BDA 1125
Linear Algebra & Linear Programming	BDA 1225
Computing for Data Science	BDA 1325
Database Management	BDA 1425
Research Methodology	BDA 1525
Foundation of data Science (programming for big Data)	BDA 2125
Advanced Analytics	BDA 2225
Machine Learning I	BDA 2325
Enabling Technologies for Data Science I	BDA 2425
Data Mining	BDA 2625
Digital Information Processing	BDA 2725
Modeling in Operations Management	BDA 3125
Enabling Technologies for Data Science II	BDA 3225
Deep learning	BDA 3325
Data Analytics on Cloud	BDA 3425
Introduction to Econometrics and Finance	BDA 3525

GENERIC ELECTIVE COURSES (GSE)/ Can include open electives offered	
Course Title	Code Number
Value thinking	BDA 2525

SKILL ENHANCEMENT COURSE (SEC) – **Any practical oriented and software-based courses offered by departments to be listed below**	
Course Title	Code Number
Basic Statistical Methods Lab	BD1P1
Linear Algebra & Linear Programming Lab	BD1P2
Computing for Data Science Lab	BD1P3
Database Management System Lab	BD1P4
Foundation of data Science (programming for big Data) Lab	BD2P1
Advanced Analytics Lab	BD2P2
Data Mining and Machine Learning Lab	BD2P3
Enabling Technologies for Data Science I Lab	BD2P4
Digital Information Processing	BD2P5
Modeling in Operation Management Lab	BD3P1
Enabling Technologies for Data Science II Lab	BD3P2
Deep learning Lab	BD3P3
Research Oriented Project/ Paper	BD3P7

VALUE ADDED COURSES (VAC) **Certificate courses that add value to the core papers can be listed.**	
Course Title	Code Number
Outreach	
Ignitors	

Program Objectives (POs) — M.Sc. in Big Data Analytics

PO1 — Domain Expertise & Applied Analytics
Apply advanced mathematical, statistical, machine-learning and deep learning methods to formulate, implement and evaluate data-driven solutions across various domains using tools and platforms.

PO2 — Big-Data Engineering & Platforms
Design, build and manage scalable data pipelines and analytics solutions using contemporary big-data technologies and best data engineering practices.

PO3 — Programming & Reproducible Practice
Demonstrate proficiency in programming and software practices to develop reproducible data-science workflows, automated experiments, and production-ready components for deployment.

PO4 — Research, Critical Thinking & Innovation
Conduct rigorous applied research or industry R&D—formulating hypotheses, selecting appropriate techniques, validating models, and communicating findings in scholarly and industry formats.

PO5 — Problem Solving & Decision Support
Analyse and abstract organizational problems into analytical formulations, select suitable modelling approaches and deliver actionable decision support using quantitative evidence.

PO6 — Ethical, Legal & Social Responsibility
Practice data science responsibly—addressing data privacy, model fairness, reproducibility, copyright/IPR, and the societal impacts of analytics—consistent with research ethics and professional norms taught in the program.

PO7 — Communication & Stakeholder Engagement
Clear communication of technical concepts, model assumptions, limitations and business implications to multidisciplinary stakeholders using visualizations, reports, dashboards and presentations.

PO8 — Teamwork & Project Management
Collaborate effectively within multidisciplinary teams to plan, execute, and manage analytics projects from conception to deployment.

PO9 — Lifelong Learning & Professional Growth
Continually update their competencies to adapt to rapidly evolving tools, methods and industry needs—pursuing professional certifications, research, or entrepreneurial initiatives in analytics and adjacent fields.

PO10 — Societal & Domain Impact Orientation
Apply data analytics to develop sustainable, equitable and value-creating solutions that address real-world challenges reflecting the program's mission to serve society.

Course Outcomes and Course Contents

Semester	FIRST
Paper Code	BDA1125
Paper Title	STATISTICAL AND PROBABILISTIC METHODS
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The course aims to explain the major concepts used in statistical and probabilistic techniques. The objective is to develop analytical ability to solve real-world problems that involve the interplay of statistical and probabilistic thinking.

COURSE OUTCOMES:

CO1: Understand the concept of data collection and analysis.
CO2: Evolve effective data visualization and dashboarding techniques that facilitate thought processes, provide new and rich insights, and promote interactivity.
CO3: Knowledge of analytical techniques with particular focus on predictive methods and probabilistic thinking
CO4: Discuss how these foundations will lead to novel approaches in big data management and analytical intelligence.

UNIT 1: DATA COLLECTION **2 Hrs.**

Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues, cleaning and treatment of missing data, sampling techniques.

UNIT 2: DATA VISUALIZATION **3 Hrs.**

Principles of data visualization and different methods of interacting with business data via chatbots and dashboards

UNIT 3: ESSENTIAL STATISTICAL METRICS **3 Hrs.**

Measures of central tendency and dispersion, covariance, correlations, regression, non-parametric methods, Measures of Skewness and Kurtosis.

UNIT 4: PROBABILITY FOUNDATIONS **10 Hrs.**

Sets and events, idea of sample space, probability of union, intersection and complementary events, conditional probability, independence, Bayes Theorem, and Bayesian approaches.

UNIT 5: PROBABILITY DISTRIBUTIONS **17 Hrs**

Random variables and their expectation and variance. Discrete and continuous probability distributions like Binomial, Poisson, Hypergeometric, Normal, Exponential, Chi square, t and F. Central Limit Theorem, Two types of errors, test statistic, parametric and non-parametric tests for equality of means and variances

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Statistical Inference : P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall
2. Introduction to Linear Regression Analysis: Douglas C. Montgomery
3. Introduction to Mathematical Statistics, Robert V. Hogg, Joseph W. McKean, Allen T. Craig, Pearson
4. An Introduction to Probability and Statistics, Vijay K. Rohatgi and K. Md. Ehsanes Saleh

Semester	FIRST
Paper Code	BDA1225
Paper Title	LINEAR ALGEBRA & LINEAR PROGRAMMING
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

To help students understand the 'intuition' behind the concepts of Linear Algebra and which in turn will help them to see its applications in later courses.

COURSE OUTCOMES:

CO1: Understand the most fundamental concept 'vector' that constructs Linear Algebra.
CO2: Able to gain knowledge of two Fundamental topics of Linear Algebra and Vector Space
CO3: Understanding two Fundamentals topics of Linear Algebra and Linear Transformation
CO4: Building the Basics of Linear Programming

UNIT 1: VECTORS **12 Hrs**

Introduction to Linear Algebra, Difference Between Linear Algebra & Matrix Analysis, Revision of Basic Geometry, Definition of Vectors - Examples, Two Fundamental Vectors – Geometric Vectors and R_n Vectors, Properties of Vectors, Linear Combination of Vectors, Decomposition of Vectors, Linear Independent & Linearly Dependent Vectors and Span of Vectors.

UNIT 2: VECTOR SPACE **10 Hrs**

Definition of Vector Space – Examples, Definition of Subspaces – Examples, Union & Intersection of Subspaces, Definition of Basis Vectors – Standard Basis and Dimension of Vector Space

UNIT 3: LINEAR TRANSFORMATION **10 Hrs**

Definition of Linear Transformation – Examples, Introduction to Matrix, Matrix as Linear Transformation, Matrix Multiplication (Composition of Linear Transformations) – Three Perspectives: 1. Column, 2. Row & 3. Dot Product, Concept of Determinant – Area, Volume, Hyper-plane, etc., System of Linear Equations – Column & Null Space, Gaussian Elimination, Row Reduced Echelon Form, Eigenvalues & Eigenvectors, Inverse Matrix and Positive Definite & Semi-Definite Matrix.

UNIT 4: LINEAR PROGRAMMING **8 Hrs.**

Introduction to Linear Programming – Examples, Problems in LP, Convex Sets, Corner Points, Feasibility, Basic Feasible Solutions and Simplex Method

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Introduction to Linear Algebra, Gilbert Strang 5th Edition.
2. Linear Programming, G. Hadley.

Semester	FIRST
Paper Code	BDA1325
Paper Title	COMPUTING FOR DATA SCIENCE
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

This course introduces and equips the student with computing techniques which enables implementing data science processes at ease. It will help students build strong fundamentals in computing and programming methodologies.

COURSE OUTCOMES:

CO1: Learning R and its purpose/usage.
CO2: Learn algorithms and lay strong programming foundations and skills.
CO3: Understanding mathematics & its challenges in computations.
CO4: Understanding simulation techniques useful for running simulated experiments.

UNIT 1: R PROGRAMMING **6 Hrs**.

Introduction to R Programming, its usage and illustrations.

UNIT 2: CONCEPTS OF COMPUTATION: ALGORITHMS **13 Hrs**.

Design and Algorithms, Convergence, Complexity with illustrations, Linear and Binary Search, Sorting Techniques (Bubble, Insertion, Selection, Quick, Merge, Heap) and Memory Handing Strategies.

UNIT 3: CONCEPTS OF COMPUTATION: NUMERICAL METHODS **13 Hrs**.

Introduction to Numerical Methods, examples: Newton-Raphson, Steepest Ascent, etc. Problem solving sessions and self-study.

UNIT 4: COMPUTING METHODOLOGIES **8 Hrs.**

Introduction to Simulations, Monte-Carlo Simulations, Statistical Models in Simulations, Random Number Generators,

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Computer Algorithms, Ellis Horowitz
2. Discrete-Event System Simulation, Jerry Banks

Semester	First
Paper Code	BDA1425
Paper Title	DATABASE MANAGEMENT SYSTEM
Number of teaching hrs per week	3 Hrs

Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

This course concentrates on introduction, principles, design and implementation of DBMS. It introduces about the distributed system and brief about data mining and data warehouse. To provide strong foundation of database concepts and develop skills for the design and implementation of a database application with a brief exposure to advanced database concepts.

COURSE OUTCOMES:

CO1: Understanding the fundamental concepts of Database Management systems
CO2: Understanding the concepts of Database models.
CO3: Understanding the core terms, concepts, and tools of relational database management systems.
CO4: Understanding database design and logic development for database programming.

UNIT 1: DATABASE MANAGEMENT SYSTEM INTRODUCTION 10 Hrs.

Data- Database- Database management system- Characteristics of the database approach- Role of Database administrators- Role of Database Designers- End Users- Advantages of Using a DBMS-Data models, Schema and Instances –Database design - Database Engine – 1 tier architecture – 2 tier architecture- 3 tier architecture – History of Database Management systems- Types of Databases.

UNIT 2: DATABASE MODELS AND IMPLEMENTATION 10 Hrs.

Data Model and Types of Data Model- Relational Data Model- Hierarchical Model- Network Data Model- Object/Relational Model- Object-Oriented Model- Entity-Relationship Model- Modeling using E-R Diagrams- Notation used in E-R Model- Relationships and Relationship Types- Cardinalities. Subclasses, Super classes and Inheritance – Specialization and Generalization – Characteristics of Specialization and Generalization – Modeling of UNION types with categories- An example University EER Schema.

UNIT 3: RELATIONAL DATABASES 10 Hrs.

Structure of relational databases- Properties of relational databases and Tables –Structure of relational databases – Database Schema – Armstrong Axioms – Functional Dependency-Anomalies in a Database- Properties of Normalized Relations- First Normalization- Second Normal Form Relation- Third Normal Form- Boyce-Codd Normal Form (BNCF).

UNIT 4: SQL AND ADDITIONAL CONCEPTS 10 Hrs.

Categories of SQL Commands; Data Definition; Data Manipulation Statements, SELECT - The Basic Form, Subqueries, Functions, GROUP BY Feature, Updating the Database, Data Definition Facilities. MongoDB Overview- MongoDB Data modeling.

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Elmasri Ramez and Navathe Shamkant B, Fundamentals of Database Systems, Addison-Wesley, 6th Edition, 2010.
2. Silberschatz, Korth, Sudarshan, Database System Concepts, 5 Edition, McGraw Hill, 2006.
3. O`neil Patricand, O`neil Elizabeth, Database Principles, Programming and Performance, 2nd Edition, Margon Kaufmann Publishers Inc, 2008.

Semester	FIRST
Paper Code	BDA1525
Paper Title	RESEARCH METHODOLOGY
Number of teaching hrs per week	2 Hrs
Total number of teaching hrs per semester	30
Number of credits	2

COURSE OBJECTIVES:

The main objectives of this paper are to develop the research aptitude among the students, to make them familiar with different research methods and techniques, to create an understanding of the basic orientation, philosophy and methods of research enquiry. Participation in the course will enable the students to become more sensitized to the social and psychological realities of individual- society and take a creative and ethical approach to extending the knowledge base to the world of practice.

COURSE OUTCOMES:

After completion of the course, students should be able to:

CO1: Understand the meaning and importance of research
CO2: Understand the concept of research design and survey methodology
CO3: Collection of data, processing of data and descriptive measures of data
CO4: Inferential analysis of data with hypothesis testing and multivariate techniques

Unit I: Concept of Research 5 hrs

Meaning of research, Objectives of research, Types of research, Research approaches, Significance of research, Research methods versus methodology, Research and scientific methods, Research processes, Criteria for good research, Research problem, Selecting the problem, Necessity of defining the problem,

Techniques involved in defining a problem, importance of literature review in defining a problem, literature review-primary and secondary sources, reviews, monograph, patents, research databases, web as a source, searching the web, critical literature review, identifying gap areas from literature and research database, development of working hypothesis.

UNIT II: Research Design and Sample Surveys 7 hrs

Meaning and need for research design, features of a good design. Important concepts relating to research design: Dependent and independent variables, extraneous variables, Control, Research hypothesis, Experimental and non-experimental hypothesis –testing research, Experimental and control group. Different research designs: Research design in case of exploratory research studies, Research design in case of hypothesis- testing, research studies, basic principles of experimental designs, Important Experimental Designs, Sampling Design, steps in sample design, criteria of selecting a sampling procedure, characteristics of a good sample design, different types of sample design.

UNIT III: Data Collection and Data Processing 8 hrs

Measurements in Research, Measurement Scales, Sources of errors in measurement, Collection of primary data: Observation Method, Interview Method, through questionnaires, through schedules, difference between questionnaire and schedule, Collection of secondary data, Selection of appropriate methods for data collection, Case study method. Analysis using statistical techniques: Inferential statistics analysis, Association statistics analysis, Casual analysis, t-test, ANOVA, Correlation and Multiple regression, Chi-square test.

UNIT IV: Report Writing & Research Ethics, IPR and Publishing 10 hrs

Locating Information on a Topic of Interest, Acquiring Copies of Articles of Interest The Nature of Scientific Variables, Conceptual Versus Operational Definitions of Variables, Levels of Measurement, Various Paradigms including Formism, Mechanism, Organicism, Pragmatism, The Basic Format for a Research Report, Identification of the Parts of a Research Report, Citation and Referencing Styles, Essentials of Report Writing, Aids for Writing Good Research Report. Ethics-ethical issues, ethical committees (human & animal); IPR- intellectual property rights and patent law, commercialization, copy right, royalty, trade related aspects of intellectual property rights (TRIPS); publishing- IMRAD concept and design of research paper, citation and acknowledgement, plagiarism, reproducibility and accountability.

Text Books:

1. C. R. Kothari, Research Methodology - Methods and Techniques, New Age International Publishers, 2023.
2. Mr. Ramakrishna Chintakunta, A Textbook of Intellectual Property Rights, Blue Hill, Publications, 2022

Reference Books:

1. Michael Alley, The Craft of Scientific Writing (3rd Edition), Springer, New York, 1996

2. Philip Reubens (General editor), Science and Technical Writing – A Manual of Style (2nd Edition), Routledge, New York, 2001

Code number and Title of the paper: BD1P1 STATISTICAL AND PROBABILISTIC METHODS LAB

List of Programs -

1. Diagrammatic & Graphical Representation of Data using Excel
2. Introduction to R Software
3. Measures of Central Tendency
4. Measures of Dispersion
5. Measures of Skewness and Kurtosis
6. Analysis of Univariate Data
7. Analysis of Bivariate Data
8. Fitting Probability Distributions – 1
9. Analysis of Qualitative Data
10. Project work

Course Objective

To develop practical skills in statistical analysis and probability applications using tools such as Excel and R. The course enables students to summarize, visualize, and interpret data effectively for analytical decision-making.

Course Outcomes

On completion, students will be able to:

- **CO1:** Represent and visualize data using suitable diagrammatic and graphical methods.
- **CO2:** Apply R software for performing basic statistical computations and analysis.
- **CO3:** Compute and interpret measures of central tendency, dispersion, skewness, and kurtosis.
- **CO4:** Analyse univariate and bivariate data and fit suitable probability distributions.
- **CO5:** Conduct data-driven mini projects demonstrating applied statistical reasoning and reporting.

Code number and Title of the paper : BD1P2 LINEAR ALGEBRA & LINEAR PROGRAMMING LAB

Different Visualisation modules of Python

Code number and Title of the paper : BD1P3 COMPUTING FOR DATA SCIENCES LAB

1. Sorting algorithms
2. Searching algorithms
3. Numerical methods
4. Monte carlo simulation

Code number and Title of the paper : BD1P4 DATABASE MANAGEMENT LAB

DDL

1. EER diagram
2. DML
3. Different types of JOIN operations
4. Manipulating database using Python
5. MongoDB
6. Project

SEMESTER II

Semester	SECOND
Paper Code	BDA 2125
Paper Title	FOUNDATION OF DATA SCIENCE
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVE:

The course will introduce students to the data scientist toolkit and the underlying core concepts. It will cover the full technical pipeline from data collection (sampling methods, crawling) to processing and basic notions of statistical analysis and visualization. The module will also include advanced topics in high-performance computing.

COURSE OUTCOME:

CO1: To understand and apply the fundamental concepts in graph for solving practical problems.
CO2: Learn the fundamentals of data analytics and the data science pipeline
CO3: Learn how to scope the resources required for a data science project and understand the advanced concepts of data Science methods.
CO4: Know what analyses are possible given a particular data set, including both the state of the art of the field and inherent limitations

UNIT 1: GRAPH THEORY **10 Hrs.**

Basic Concepts, Algorithms for connectedness, Shortest path, Minimum Spanning Tree

UNIT 2: HIGH DIMENSIONAL SPACE **10 Hrs**.

Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications.

UNIT 3: RANDOM GRAPHS AND SINGULAR VALUE DECOMPOSITION (SVD)

10 Hrs.

Large graphs, G (n, p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.
SVD: Best rank k approximation, Power method for computing the SVD, Applications.

UNIT 4: RANDOM WALKS AND ALGORITHM FOR MASSIVE DATA PROBLEMS

10 Hrs.

Reflection Principle, Long leads, Changes of Sign, Illustrations. Frequency Moments of data streams, matrix algorithms.

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Foundations of Data Science: John Hopcroft & Ravindran Kannan.

Semester	SECOND
Paper Code	BDA 2225
Paper Title	ADVANCED ANALYTICS
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The course aims to explain advanced concepts and techniques in multivariate statistics, and develop the ability to solve more nuanced real-world problems requiring greater analytical intelligence.

COURSE OUTCOMES:

CO1: Understand concepts of estimation, and hypothesis testing involving multiple variables as in ANOVA
CO2: Gain expertise in powerful multivariate techniques such as principal component analysis, clustering and classification
CO3: Learn how to develop learning and prediction models that involve uncertainty.

.

UNIT 1: MULTIVARIATE TESTING AND ESTIMATION 10 Hrs.

Unbiasedness, consistency, maximum likelihood estimates, expectation maximization and bootstrap algorithms, Gauss Markov models and least square estimators, multiple linear and polynomial regression, analysis of variance.

UNIT 2: MULTIVARIATE ANALYSIS TECHNIQUES 22 Hrs.

Principal component analysis, similarity measures and clustering algorithms, classification with logistic regression and other techniques

UNIT 3: INTRODUCTION TO TIME SERIES 10 Hrs.

Components of time series, smoothing auto correlation, stationarity, ARIMA models and its variants with illustrations

UNIT 4: STOCHASTIC PROCESS 3 Hrs.

Markov Chains, classification of states, stationery distribution, idea of stochastic process.

SUGGESTED BOOKS:

1. Introduction to Linear Regression Analysis: Douglas C. Montgomery
2. Applied Multivariate Statistical Analysis : Richard A. Johnson and Dean W. Wichern, Prentice Hall, 2002
3. **S**tatistical Inference: P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall.

Semester	SECOND
Paper Code	BDA2325
Paper Title	MACHINE LEARNING I
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

This course will provide the students to understand the concepts of Machine Learning, supervised learning and their applications, the concepts and algorithms of unsupervised learning the concepts and algorithms of advanced learning.

COURSE OUTCOMES:

CO1: Design a learning model appropriate to the application.
CO2: Design a supervised learning for an application of your choice.
CO3: Design an unsupervised learning for an application of your choice.
CO4: Identify applications dimensionality reduction suitable for different types of Machine Learning with suitable justification.

UNIT 1: MACHINE LEARNING INTRODUCTION **10 Hrs**

Machine Learning–Types of Machine Learning –Machine Learning process- preliminaries, testing Machine Learning algorithms, turning data into Probabilities, and Statistics for Machine Learning Probability theory – Probability Distributions – Decision Theory.

UNIT 2: SUPERVISED LEARNING **10 Hrs**

Linear Models for Regression, Linear Models for Classification, Discriminant Functions, Probabilistic Generative Models, Probabilistic Discriminative Models, Decision Tree Learning, Bayesian Learning, Naïve Bayes, Ensemble Methods – Bagging and Boosting, Mixture of experts, Support Vector Machines. Clustering- K-means – EM Algorithm- Mixtures of Gaussians –Estimating means of K Gaussians.

UNIT 3: DIMENSIONALITY REDUCTION **8 Hrs**

Dimensionality Reduction, Linear Discriminant Analysis, Factor Analysis, Principal Components Analysis, Independent Components Analysis, TSNE.

UNIT 4: NATURAL LANGUAGE PROCESSING **12 Hrs**

Introduction to NLP - Overview of NLP, history and applications, Challenges in NLP, Text preprocessing techniques (tokenization, cleaning, normalization), Word representations (one-hot encoding, word embeddings). Lexical Analysis - Tokenization, stemming, lemmatization, Part-of-speech tagging (POS tagging), Named entity recognition (NER) techniques. Statistical Language modeling- N-grams language models, Markov chains and Hidden Markov Models (HMMs), Smoothing techniques for language models.

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Tom Mitchell, “Machine Learning”, McGraw-Hill, 1997.
2. Christopher Bishop, “Pattern Recognition and Machine Learning” Springer, 2007. \

3. Stephen Marsland, “Machine Learning – An Algorithmic Perspective”, Chapman and Hall, CRC Press, Second Edition, 2014.
4. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012.
5. Ethem Alpaydin, “Introduction to Machine Learning”, MIT Press, Third Edition, 2014.

Semester	THIRD
Paper Code	BDA2425
Paper Title	ENABLING TECHNOLOGIES FOR DATA SCIENCE I
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

This course explains the key issues in big data management and trains the students to have skills that will help them to solve complex real-world problems for prediction and decision making using different tools.

COURSE OUTCOMES:

CO1: Understand the key issues in big data management and its associated applications using Hadoop
CO2: Acquire fundamental enabling techniques and scalable algorithms like Map Reduce
CO3: Interpret business models and scientific computing paradigms, and apply software tools like HIVE for big data analytics
CO4: Achieve adequate perspectives of big data analytics in various applications using SQOOP
CO5: Gain knowledge of PIG based on Big Data applications

UNIT 1: BIG DATA AND HADOOP **12 Hrs.**

Hadoop architecture, Hadoop Versioning and configuration, Single node & Multi-node Hadoop, Hadoop commands, Models in Hadoop, Hadoop daemon, Task instance, Illustrations.

UNIT 2: MAP-REDUCE **12 Hrs.**

Framework, Developing Map-Reduce program, Life cycle method, Serialization, Running Map-Reduce in local and pseudo-distributed mode, Illustrations

UNIT 3: HIVE **6 Hrs.**

Installation, data types and commands, Illustrations.

UNIT 4: SQOOP **5 Hrs.**

Installation, Importing data, Exporting data, Running, Illustrations

UNIT 5: PIG **5 Hrs.**

Installation, Schema, Commands, Illustrations.

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Hadoop in Action : Chuck Lam, 2010, ISBN : 9781935182191
2. Data-intensive Text Processing with Map Reduce : Jimmy Lin and Chris Dyer, Morgan& Claypool Publishers, 2010

Semester	SECOND
Paper Code	BDA 2525
Paper Title	VALUE THINKING
Number of teaching hrs per week	2 Hrs
Total number of teaching hrs per semester	30
Number of credits	2

COURSE OBJECTIVES:

The course aims to improve the argumentative logic and to inculcate logical thinking. Students will understand the importance of value based living. They will gain deeper understanding abo ut the purpose of their life. They will not only understand, they will start applying the essential steps to become good leaders and value based professionals.

COURSE OUTCOMES:

CO1: Demonstrate an enhanced ability to employ evidence/information in conducting a comprehensive analysis of an issue or problem
CO2: Demonstrate an enhanced ability to draw logical conclusions and implications from the analysis of an issue or problem

Movies:

1. Twelve Angry Men
2. Roshoman by Kurosawa
3. Trial of Nuremberg

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

Case Studies:

Semester	SECOND
Paper Code	BDADE 2625
Paper Title	DATA MINING
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The course is intended to cover the comprehensive understanding of data mining concepts, techniques, and tools. Students will develop the ability to analyze large datasets and extract meaningful patterns.

COURSE OUTCOMES:

CO1: Understand and apply fundamental data mining concepts and techniques.
CO2: Use data mining tools and software to analyze large datasets.
CO3: Develop, implement, and evaluate classification, clustering, and association algorithms on real-world datasets
CO4: Apply advanced data mining techniques such as text mining, sentiment analysis, and anomaly detection.
CO5: Explore current trends and applications of data mining in various industries like e-commerce, healthcare, and social networks.

UNIT 1: INTRODUCTION **10 Hrs**

Overview of data mining, process & methodologies, types of data & data preprocessing, Modalities of data - structured data, unstructured data, semi-structured data, temporal data, spatial data, multimedia data, and streaming data. Basic concepts of data warehousing.

UNIT 2: CLASSIFICATION AND PREDICTION **10 Hrs**

Classification techniques, decision trees & rule-based classification, bayesian classification, regression analysis, evaluation of classification models.

UNIT 3: CLUSTERING AND ASSOCIATION ANALYSIS **10 Hrs**

K-means, hierarchical clustering, density-based clustering techniques, association rule mining, market basket analysis, apriori algorithm.

UNIT 4: ADVANCED DATA MINING TECHNIQUES **10 Hrs**

Text mining & sentiment analysis, web mining & social network analysis, anomaly detection, time series & sequence mining. Trends in Data Mining - Data mining in e-commerce & healthcare, big data & scalable data mining, ethical issues & privacy concerns, future trends in data mining.

UNIT 5: SELF STUDY **5 Hrs**

REFERENCE BOOKS

1. Jiawei Han, Micheline Kamber, and Jian Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.
2. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to Data Mining, Pearson, 2005.
3. Ian H. Witten, Eibe Frank, and Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2011.
4. Jure Leskovec, Anand Rajaraman, and Jeffrey Ullman, Mining of Massive Datasets, Cambridge University Press, 2014.
5. Nathan Marz and James Warren, Big Data: Principles and Best Practices of Scalable Real-Time Data Systems, Manning Publications, 2015.

Semester	SECOND
Paper Code	BDADE 2725
Paper Title	DIGITAL INFORMATION PROCESSING
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE DESCRIPTION:

The course is intended to cover Digital Image Processing techniques and its applications.

COURSE OBJECTIVES:

1. Signals and systems & Transformation techniques
2. Discrete Fourier Transform Computation
3. To become familiar with digital image fundamentals
4. To get exposed to simple image enhancement techniques in Spatial and Frequency domain.
5. To study the image segmentation and representation techniques.
6. To become familiar with image compression and recognition methods

COURSE OUTCOMES:

CO1: Ability to acquire knowledge on Signals and systems & their mathematical representation.
CO2: Ability to analyze the transformation techniques & their computation.
CO3: Know and understand the basics and fundamentals of digital image processing, such as digitization, sampling, quantization, and 2D-transforms.
CO4: Operate on images using the techniques of smoothing, sharpening and enhancement.
CO5: Learn the basics of segmentation, features extraction and compression

UNIT 1: INTRODUCTION TO SIGNALS 9 Hrs

Classification of systems: Continuous, discrete, linear, causal, stability, dynamic, recursive, time variance; classification of signals: continuous and discrete, energy and power; mathematical representation of signals; spectral density; sampling techniques, quantization, quantization error, Nyquist rate, aliasing effect

UNIT 2: DISCRETE FOURIER TRANSFORM & COMPUTATION 9 Hrs

Discrete Fourier Transform- properties, magnitude and phase representation – Computation of DFT using FFT algorithm – DIT &DIF using radix 2 FFT – Butterfly structure.

UNIT 3: FUNDAMENTALS OF IMAGE PROCESSING 9 Hrs

Introduction – Steps in Image Processing Systems – Image Acquisition – Sampling and Quantization – Pixel Relationships – Color Fundamentals and Models, File Formats, Image operations – Arithmetic, Geometric and Morphological.

UNIT 4: IMAGE ENHANCEMENT 9 Hrs

Spatial Domain Gray level Transformations Histogram Processing Spatial Filtering – Smoothing and Sharpening. Frequency Domain: Filtering in Frequency Domain – DFT, FFT, DCT – Smoothing and Sharpening filters – Homomorphic Filtering.

UNIT 5: IMAGE COMPRESSION & SEGMENTATION 9 Hrs

Detection of Discontinuities – Edge Operators – Edge Linking and Boundary Detection – Thresholding – Region Based Segmentation – Morphological WaterSheds – Motion

Segmentation. Image Compression : Fundamentals – Models – Elements of Information Theory –Error Free Compression – Lossy Compression – Compression Standards.

REFERENCE BOOKS

1. J.G. Proakis and D.G. Manolakis, 'Digital Signal Processing Principles, Algorithms and Applications', Pearson Education, New Delhi, PHI. 2003.
2. S.K. Mitra, 'Digital Signal Processing – A Computer Based Approach', McGraw Hill Edu, 2013.
3. Lonnie C.Ludeman ,"Fundamentals of Digital Signal Processing",Wiley,2013
4. Rafael C.Gonzalez and Richard E.Woods, "Digital Image Processing" Second Edition, Pearson Education, 2003
5. Milan Sonka, Vaclav Hlavac and Roger Boyle, "Image Processing, Analysis and Machine Vision", Second Edition, Thomson Learning, 2001
6. Anil K.Jain, "Fundamentals of Digital Image Processing", PHI, 2006.

LABORATORY

Code number and Title of the paper : BD2P1 Foundation of Data Science LAB

1. Shortest path algorithms(python)
2. Minimum cost algorithm(python)
3. Similarity algorithms(python)
4. G(n,p) model (using graph database)
5. SVD
6. Data stream

Code number and Title of the paper : BD2P2 Advanced Analytics LAB

List of programs –

1. Maximum Likelihood Method of Estimation.
2. Tests of Significance – 1 Formulation of Hypotheses and Types of Errors.
3. Tests of Significance – 2 Tests Concerning Single Population Mean.
4. Tests of Significance – 3 Tests Concerning Two Populations Mean.
5. Tests of Significance – 4 Tests Concerning Population Variance.
6. ONE WAY ANOVA.
7. TWO WAY ANOVA.
8. Applied Regression Analysis
9. Logistic Regression.
10. Project

Code number and Title of the paper : BD2P3 Datamining and Machine Learning LAB

Data Mining

1. Handling missing values, data normalization, transformation, and Identifying outliers in datasets
2. Conducting Market Basket Analysis using the Apriori algorithm
3. Forecasting future trends using time series data.
4. Scraping and analyzing data from web sources.

Machine Learning

1. Linear Regression, Logistic Regression
2. Naïve Bayes, Decision Tree
3. Support Vector Machines
4. Comparing model performances using precision, recall, F1-score, and ROC curves.
5. Unsupervised Learning(K-means), hierarchical clustering
6. Dimensionality Reduction
7. Extracting and analyzing sentiment from textual data

Code number and Title of the paper : BD2P4 Enabling Technologies for Data Science I LAB

1. Hadoop
2. Map-Reduce
3. HIVE
4. SQOOP
5. PIG

Projects using R and/or Python(Include Power BI)

SEMESTER III

Semester	THIRD
Paper Code	BDA 3125
Paper Title	MODELING IN OPERATIONS MANAGEMENT
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

This course concentrates on the skills to build their own model formulations, to expand existing model formulations, to critically evaluate the impact of model assumptions and to choose an appropriate solution technique for a given model formulation. This will develop analytical ability to solve real-world problems using these methodologies.

COURSE OUTCOMES:

CO1: Understanding concepts of venture analytics, applications, quantitative methods and its strategic frameworks

CO2: Understanding concepts of Banking analytics, applications, quantitative methods and its strategic frameworks

CO3: Understanding concepts of Marketing analytics, applications, quantitative methods and its strategic frameworks
CO4: Understanding concepts of Healthcare analytics, applications, quantitative methods and its strategic frameworks
CO5: Understanding concepts of Retail analytics, applications, quantitative methods and its strategic frameworks. Emphasis on Consumer Packaged Goods (CPG).
CO6: Understanding concepts of Supply chain analytics, applications, quantitative methods and its strategic frameworks

UNIT 1: VENTURE ANALYTICS **5 Hrs**

UNIT 2: BANKING ANALYTICS **7 Hrs**

UNIT 3: MARKETING ANALYTICS **7 Hrs**

UNIT 4: HEALTHCARE ANALYTICS **7 Hrs**

UNIT 5: RETAIL ANALYTICS WITH CPG **7 Hrs**

Introduction to key terms and KPIs in retail analytics, Overview of Data sources in retail, Customer Segmentation and Profiling, Customer Lifetime Value (CLTV), Market Basket analysis and Recommendation systems, Price Elasticity of Demand, Cannibalization Analysis, A/B Testing in Retail, Promotional Effectiveness, Market Mix Modeling, Price Pack Architecture (PPA), Trade Promotions Optimization (TPO), Demand Forecasting, Dynamic Pricing models.

UNIT 6: SUPPLY CHAIN ANALYTICS **7 Hrs**

SELF STUDY **5 Hrs**

SUGGESTED BOOKS:

1. Introduction to Mathematical Statistics, Robert V. Hogg, Joseph W. McKean, Allen T. Craig,Pearson
2. An Introduction to Probability and Statistics, Vijay K. Rohatgi and K. Md. Ehsanes Saleh
3. Introductory Econometrics , Jeffrey M . Wooldridge

Semester	THIRD
Paper Code	BDA 3225
Paper Title	ENABLING TECHNOLOGIES FOR DATA SCIENCE II

Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The students will learn the concepts of Data Warehousing and its implementations. The use of spark, Scala , Mahoot will be explained and applications will be provided.

COURSE OUTCOMES:

CO1: Basic knowledge of a Data Warehouse system
CO2: Understand data pre-processing techniques during data warehousing implementation
CO3: Learn to apply the concept of Spark
CO4: Understand the concepts of Scala and apply them
CO5: Applications of Mahoot and solving the real life problems

UNIT 1: DATA WAREHOUSING AND MODELING **5 Hrs.**

Basic Concepts: Data Warehousing: A multitier Architecture, Data warehouse models: Enterprise warehouse, Data mart and virtual warehouse, Extraction, Transformation and loading, Data Cube: A multidimensional data model, Stars, Snowflakes and Fact constellations: Schemas for multidimensional Data models, Dimensions: The role of concept Hierarchies, Measures: Their Categorization and computation, Typical OLAP Operations.

UNIT 2: DATA WAREHOUSE IMPLEMENTATION & DATA MINING **10 Hrs.**

Efficient Data Cube computation: An overview, Indexing OLAP Data: Bitmap index and join index, Efficient processing of OLAP Queries, OLAP server Architecture ROLAP versus MOLAP Versus HOLAP. : Introduction: What is data mining, Challenges, Data Mining Tasks, Data: Types of Data, Data Quality, Data Preprocessing, Measures of Similarity and Dissimilarity,

UNIT 3: SPARK **10 Hrs.**

UNIT 4: SCALA **10 Hrs.**

UNIT 5: MAHOUT **5 Hrs.**

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Tan P. N., Steinbach M & Kumar V. “Introduction to Data Mining” Pearson Education, 2006.

2. Prateek Bhatia, “Data Mining and Data warehousing”, Cambridge University Press, 2019.

Semester	THIRD
Paper Code	BDA 3325
Paper Title	DEEP LEARNING
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

- To have a thorough understanding of Deep learning concepts.
- To study CNN, RCNN and Fast RCNN
- To understand NLP and Genetic algorithm.
- To understand basic concepts of probabilistic graphical models and different inference techniques.

COURSE OUTCOMES:

CO1: Develop and Train Deep Neural Networks
CO2: Develop a CNN, R-CNN, Fast R-CNN for detection and recognition.
CO3: Work with NLP and Word Embedding’s.
CO4: Design probabilistic graphical models and different inference techniques.

UNIT I: DEEP LEARNING CONCEPTS 10 Hrs

Fundamentals about Deep Learning. Perception Learning Algorithms. Probabilistic modelling. Early Neural Networks. How deep learning different from Machine Learning. Scalars. Vectors. Matrixes, Higher Dimensional Tensors. Manipulating Tensors. Vector Data. Time Series Data. Image Data. Video Data. About Neural Network. Building Blocks of Neural Network. Optimizers. Activation Functions. Loss Functions. Data Pre-processing for neural networks, Feature Engineering. CNN: Linear Time Invariant. Image Processing Filtering. Building a convolutional neural network. Input Layers, Convolution Layers. Pooling Layers. Dense Layers. Backpropagation through the Convolutional Layer. Filters and Feature Maps. Backpropagation through the Pooling Layers. Dropout Layers and Regularization. Batch Normalization. Various Activation Functions. Various Optimizers. Transfer Learning with Image Data. Transfer learning.

UNIT II: RECURRENT NEURAL NETWORKS **10 Hrs**

Text Classification: Processing raw text, Categorizing and Tagging words, from text to tokens, Text Classification, Applications: Summarization, Question Answering, Architecture of RNNs, Forward and Backward Pass, Challenges with RNNs, LSTM, BERT, introduction to LLM.

UNIT III: RETRIEVAL AUGMENTED GENERATION **10 Hrs**

Retrieval-Augmented Generation (RAG) in LLM, RAG Mechanism, RAG implementation in Python. Reinforcement Learning- Overview of Reinforcement Learning, Basic Solution Methods - Dynamic Programming, Monte Carlo Methods, Temporal-Difference Learning, Comparison of Learning Techniques. Value Function Approximation - Model-Free Reinforcement Learning.

UNIT IV: GENETIC ALGORITHMS **10 Hrs**

Genetic Algorithms: Motivation- Representing hypothesis- Genetic Operators- Fitness functions & selections-Extensions- Hypothesis space search.

Graph Models: Probabilistic Graphical Models – Motivation –Foundations – Probability Theory – Graphs -Independence Properties - Bayesian Network Representation - Independence in Graphs – From Distribution to Graphs, Inference - Markov Chain Monte Carlo Methods.

SELF STUDY **5 Hrs.**

REFERENCE BOOKS

1. Tom Mitchell, “Machine Learning”, McGraw-Hill, 1997.
2. Christopher Bishop, “Pattern Recognition and Machine Learning” Springer, 2007. \
3. Deep Learning A Practitioner’s Approach Josh Patterson and Adam Gibson O’Reilly Media,Inc.2017
4. Learn Keras for Deep Neural Networks, Jojo Moolayil, Apress,2018
5. Deep Learning Projects Using TensorFlow 2, Vinita Silaparasetty, Apress, 2020
6. Deep Learning with Python, FRANÇOIS CHOLLET, MANNING SHELTER ISLAND,2017Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012.
7. Ethem Alpaydin, “Introduction to Machine Learning”, MIT Press, Third Edition, 2014.
8. D. Koller and N. Friedman, “Probabilistic Graphical Models: Principles and Techniques”, MIT Press, 2009.
9. Probabilistic Machine Learning: An Introduction by Kevin Patrick Murphy.MIT Press, March 2022.

Semester	THIRD
Paper Code	BDA 3425
Paper Title	DATA ANALYTICS ON CLOUD
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVE:

This course will expose you to the data analytics practices executed in the business world. We will explore such key areas as the analytical process, how data is created, stored, accessed, and how the organization works with data and creates the environment in which analytics can flourish. What you learn in this course will give you a strong foundation in all the areas that support analytics and will help you to better position yourself for success within your organization.

COURSE OUTCOMES:

CO1: To understand the concept of cloud and utility computing.
CO2: To understand the various issues in cloud computing.
CO3: To familiarize themselves with the lead players in cloud.
CO4: To appreciate the emergence of cloud as the next generation computing paradigm.

UNIT 1: INTRODUCTION **10 Hrs**

Introduction- Historical Development – Cloud Computing Architecture – The Cloud Reference Model – Cloud Characteristics –Cloud Deployment Models: Public, Private, Community, Hybrid Clouds- Cloud Delivery Models: IaaS, PaaS, SaaS – Open-Source Private Cloud Software: Eucalyptus, Open Nebula, Open Stack.

UNIT 2: VIRTUALIZATION **10 Hrs**

Data Center Technology – Virtualization – Characteristics of Virtualized Environments - Taxonomy of Virtualization Techniques – Virtualization and Cloud Computing –Pros and Cons of Virtualization – Implementation Levels of Virtualization – Tools and Mechanisms: Xen, VMWare, Microsoft Hyper-V, KVM, Virtual Box

UNIT 3: CLOUD COMPUTING MECHANISM **10 Hrs**

Cloud Infrastructure Mechanism: Cloud Storage, Cloud Usage Monitor, Resource Replication – Specialized Cloud Mechanism: Load Balancer, SLA Monitor, Pay-per-use Monitor, Audit Monitor, Failover System, Hypervisor, Resource Cluster, Multi Device Broker, State Management Database – Cloud Management Mechanism: Remote Administration System, Resource Management System, SLA Management System, Billing Management System

UNIT 4: SECURITY IN THE CLOUD **10 Hrs**

Basic Terms and Concepts – Threat Agents – Cloud Security Threats –Cloud Security Mechanism: Encryption, Hashing, Digital Signature, Public Key Infrastructure, Identity and Access Management, Single Sign-on, Cloud Based Security Groups, Hardened Virtual Server Images.

SELF STUDY **5 Hrs**

SUGGESTED BOOKS:

1. Thomas Erl, ZaighamMahood, Ricardo Puttini, "Cloud Computing, Concept, Technology & Architecture", Prentice Hall, 2013.
2. RajkumarBuyya, Christian Vecchiola, S. ThamaraiSelvi, "Mastering Cloud Computing", Tata McGraw-Hill, 2013.
3. Toby Velte, Anthony Velte, Robert C. Elsenpeter, "Cloud Computing, A Practical Approach",Tata McGraw-Hill Edition, 2010.
4. ArshdeepBahga, Vijay Madisetti, "Cloud Computing: A Hands-On Approach", Universities Press(India) Private Limited, 2014.
5. Tom White, "Hadoop: The Definitive Guide", O'Reilly Media, 4th Edition, 2015.
6. James E Smith and Ravi Nair, "Virtual Machines", Elsevier, 2005.
7. John Rittinghouse & James Ransome, "Cloud Computing, Implementation, Management and Strategy", CRC Press, 2010.

Semester	THIRD
Paper Code	BDA 3525
Paper Title	INTRODUCTION TO ECONOMETRICS AND FINANCE
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

To equip the students with the necessary skills, including both the acquisition of habits of thought and knowledge of the techniques of modern econometrics, required for applied research in development economics and data analytic industry.

COURSE OUTCOMES:

CO1: To understand the analysis of Panel data and apply different methods to the models.

CO2: To know the Generalised Method of Moments (GMM) and testing of the moments using the methodology.
CO3: To solve the Simultaneous equations using different methods.
CO4: To understand the concept of Cointegration using models.
CO5: Different model making and comparing the effects of these models to understand them.

UNIT 1: ANALYSIS OF PANEL DATA **14 Hrs.**

Fixed Effects Estimation, Random Effects Model, The Correlation Random Effects Approach, Applying Panel Data Methods to Other Data Structures

UNIT 2: GENERALIZED METHOD OF MOMENTS (GMM) **14 Hrs.**

GMM estimator, Two step optimal GMM estimator, Adding moment conditions, Asymptotic theory for GMM, Conditional moment restrictions, Bias in GMM, Testing in GMM, Small bias methods

UNIT 3: SIMULTANUOUS EQUATIONS SYSTEM **5 Hrs.**
Least Squares, Bias Problem, Estimation Method.

UNIT 4: COINTEGRATION **3 Hrs.**

Concept, two variable model, Engle-Granger Method, Vector autoregressions (VAR), Vector error correlation model (VECM)

UNIT 5: ARCH/GARCH/SV MODELS, SOME IMPORTANT GENERALIZATIONS LIKE EGARCH & GJR MODELS, ARCH –M MODELS. **4 Hrs.**

ARCH model, Estimation of ARCH model, GARCH model, Generalisation on Models, EJR model, GARCH model, Analysis on models

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS :

1. The Econometrics of Financial Markets : J. Campbell, A.Lo and C. Mackinlay
2. Econometric Analysis : William H. Greene

LABORATORY

BD3P1: MODELING IN OPERATION MANAGEMENT LAB

Project

BD3P2: ENABLING TECHNOLOGIES FOR DATA SCIENCE II LAB

List of Experiments

Spark

Scala
Mahout

BD3P3: DEEP LEARNING LAB

Project

BDA3P7: Research Paper oriented LAB

Research Paper