ST JOSEPH'S UNIVERSITY BANGALORE

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 Advance Computing (PG) Revision Year - 2024						
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 1124	Statistical and probabilistic methods	45	03	03	50-2Hrs
Theory	BDA 1224	Linear Algebra & Linear Programming	45	03	03	50-2Hrs
Theory	BDA 1324	Computing for Data Science	45	03	03	50-2Hrs
Theory	BDA 1424	Database Management System	45	03	03	50-2Hrs
Theory	BDA 1524	Python Programming	45	03	03	50-2Hrs
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
Practical	BD1P5	Python Lab	30	02	01	50-2Hrs
Total Number of credits:			**24**			
Semester 2	**Code Number**	**Title**	**No. of Hours of Instructio**	**Number of teaching hrs /week**	**Number of credits**	**Max marks for SE -duration of examination**

			ns			
Theory	BDA 2124	Foundation of data Science (Programming for Big Data)	45	03	03	50-2Hrs
Theory	BDA 2224	Advanced analytics	45	03	03	50-2Hrs
Theory	BDA 2324	Machine Learning I	45	03	03	50-2Hrs
Theory	BDA 2424	Enabling Technologies for Data Science I	45	03	03	50-2Hrs
Theory	BDA 2524	Value thinking	45	03	03	50-2Hrs
Theory (DE)	BDADE 2624	Natural Language Processing			03	
Theory (DE)	BDADE 2724	Digital Image Processing	45	03	03	50-2Hrs
Theory	BDADE 2821	Research Methodology	45	03	03	50-2Hrs
Practical	BD2P1	Foundation of data Science (Programming for Big Data) Lab	30	02	01	50-2Hrs
Practical	BD2P2	ADVANCED ANALYTICS Lab	30	02	01	50-2Hrs
Practical	BD2P3	Machine Learning I Lab	30	02	01	50-2Hrs
Practical	BD2P4	Enabling Technologies for Data Science I Lab	30	02	01	50-2Hrs
Practical	BD2P5	Natural Language Processing Lab	30	02	01	50-2Hrs
Practical	BD2P6	Digital Image Processing Lab	30	02	01	50-2Hrs
Total Number of credits:			**26**			
Semester 3	**Code Number**	**Title**	**No. of Hour**	**Number of teachin**	**Number of credit**	**Max marks for SE - duration of examination**

			s of Instructions	**g hrs /week**	**s**	
Theory	BDA3124	Modeling in Operations Management	45	03	03	50-2Hrs
Theory	BDA3224	Enabling Technologies for Data Science II	45	03	03	50-2Hrs
Theory	BDA 3324	Machine Learning II	45	03	03	50-2Hrs
Theory	BDA 3424	Data Analytics on Cloud	45	03	03	50-2Hrs
Theory	BDA 3524	Introduction to Econometrics and Finance	45	03	03	50-2Hrs
Theory	BDA 3624	Digital Signal Processing	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	Machine Learning II Lab	30	02	01	50-2Hrs
Theory	BD3P7	Research Oriented Project/ Paper	30	02	01	50-2Hrs
Total Number of credits:			**22**			
Semester 4	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of teaching hrs /week**	**Number of credits**	**Max marks for SE - duration of examination**
Practical	BDA4IN24	Internship/ project	750		25	100

Practical	BD4P1	Dissertation	60	4	2	100
		IGNITORS/ OUTREACH			03	
Total Number of credits:			**31**			

CORE COURSES (CC)	
Course Title	Code Number
Statistical and Probabilistic Methods	BDA 1121
~~Probability & Stochastic Process~~	~~BDA 1221~~
Linear Algebra & Linear Programming	BDA 1321
Computing for Data Science I	BDA 1421
Database Management	BDA 1521
Python Programming	BDA 1621
Foundation of data Science (programming for big Data)	BDA 2121
~~Advance Statistical Method~~	~~BDA 2221~~
Machine Learning, I	BDA 2321
Enabling Technologies for Data Science I	BDA 2421
Modeling in Operations Management	BDA3121
Enabling Technologies for Data Science II	BDA3221
Machine Learning II	BDA 3321
Cloud Computing	BDA 3421

DISCIPLINE SPECIFIC ELECTIVE COURSES (DSE)	
Course Title	Code Number
Multivariate Statistics	BDADE 2621
Introduction to Econometrics and Finance	BDADE 3521
Bioinformatics	BDADE 3621
Digital Image Processing	BDADE 2721
Digital signal Processing	BDADE 2821

GENERIC ELECTIVE COURSES (GSE)/ Can include open electives offered	
Course Title	Code Number
Value thinking	BDA 2521

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
Probability & Stochastic Process Lab	BD1P2
Linear Algebra & Linear Programming Lab	BD1P3
Computing for Data Science I Lab	BD1P4
Database Management Lab	BD1P5
Python Lab	BD1P6
Foundation of data Science (programming for big Data) Lab	BD2P1
Advance Statistical Method Lab	BD2P2
Machine Learning, I lab	BD2P3
Enabling Technologies for Data Science I Lab	BD2P4
Multivariate Statistics Lab	BD2P5
Modeling in Operation Management Lab	BD3P1
Enabling Technologies for Data Science II Lab	BD3P2
Machine Learning II Lab	BD3P3
Cloud Platform	BD3P4
Introduction to Econometrics and Finance Lab	BD3P5
Bioinformatics Lab	BD3P6

Online courses offered or recommended by the department to be listed	
Course Title	Code Number
Natural Language Processing	
Computer Vision	

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

Outreach	
Ignitors	

Course Outcomes and Course Contents

Semester	FIRST
Paper Code	BDA1124
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	BDA1224
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	BDA1324
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	BDA1424
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	BDA1524
Paper Title	Python Programming
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVE:

This Python Programming course leads students from the basics of writing and running Python scripts to more advanced features such as file operations, regular expressions, working with binary data, and using the extensive functionality of Python modules. Extra emphasis is placed on features unique to Python, such as Data Handling and Visualization.

COURSE OUTCOMES:

CO1: Understand the basic concepts and principles of Python programming and able to implement various control statements.

CO2: Able to gain insight knowledge towards Functions, I/O, File Handling and Packages.

CO3: Gain knowledge of object-oriented programming in Python.

CO4: Get the knowledge of various Data Handling mechanism in Python using NumPy.

CO5: Understand how to handle the data with Visualization models.

UNIT 1: INTRODUCTION TO PYTHON INTERPRETER 8 Hrs.

Python - Introduction, Advantages and Disadvantages, History, Features, Applications, Python Internals, Runtime Structure, Basic Syntax, Python Identifiers, Reserved Keywords, Data Types, List, Tuple, Dictionary, Set.

Control statements

while loop, for loop, if statement, break statement, continue statement

UNIT 2: FUNCTIONS, I/O, FILE HANDLING, PACKAGES/LIBRARIES 8 Hrs.

Functions - Define, call, pass by reference, Function Arguments, Anonymous Function or Lambda Function, return statement.

I/O - Handling Files, Types of Files, Open(), close(), Different modes, Read & Write, file positions, File Seek, OS File/Directory Methods - Types and Methods

Packages/Libraries - Modules, import statement, packages.

UNIT 3: EXCEPTION HANDLING, OO PROGRAMMING **8 Hrs.**

Exception Handling - Exception Types, Handling Exceptions, Raising Exceptions

OO Programming - Classes, Objects, creating object, self-parameter, init function, destructors, privacy in python, Inheritance and its types, Polymorphism - Method overloading, method overriding, constructor overriding, operator overloading.

UNIT 4: PYTHON FOR DATA HANDLING **8 Hrs.**

Basics of Numpy arrays aggregations computations on arrays comparisons, masks, boolean logic fancy indexing structured arrays Data manipulation with Pandas data indexing and selection operating on data missing data hierarchical indexing combining datasets

aggregation and grouping pivot tables

UNIT 5: PYTHON FOR DATA VISUALIZATION **8 Hrs.**

Visualization with matplotlib line plots scatter plots visualizing errors density and contour plots histograms, binnings, and density three-dimensional plotting geographic data analysis using statmodels and seaborn graph plotting using Plotly interactive data visualization using Bokeh.

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Python in easy steps - Mike McGrath, In Easy Steps Limited, Second Edition

2. “Hello World” - Computer Programming for Kids and other Beginners - Warren and Carter, Manning Publications, 2014

3. Jake VanderPlas, “Python Data Science Handbook”, O’Reilly, 2016. (Parts of chapters 2 4 for Units IV and V)

4. Python3 Tutorial - Tutorialspoint

5. Allen B. Downey, “Think Stats: Exploratory Data Analysis in Python”, Green Tea Press, 2014.

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

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

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

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

Code number and Title of the paper : BD1P5 PYTHON PROGRAMMING LAB

List of programs –

1. Introduction to Python interpreter
2. Control statements
3. functions, I/O, File handling, Packages/Libraries
4. Exception Handling, OO Programming.

SEMESTER II

Semester	SECOND
Paper Code	BDA 2124
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 & RavindranKannan.

Semester	SECOND
Paper Code	BDA 2224
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 **5 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.

SELF STUDY **5 Hrs.**

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	BDA2324
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.

UNIT 3: UNSUPERVISED LEARNING **10 Hrs.**

Clustering- K-means – EM Algorithm- Mixtures of Gaussians –Estimating means of K Gaussians

UNIT 4: DIMENSIONALITY REDUCTION **10 Hrs.**

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

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 andHall, 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	BDA2424
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 2524
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 2624
Paper Title	NLP
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The main objectives of this paper is to provide a comprehensive view of building real-world natural language processing (NLP) applications. The goal of this course is to provide an overview of NLP foundations. The course covers the most important topics in the field, keeping a balance between theory and practice. This course focuses on a compact set of methods unified by the concepts of learning and search, which can solve a remarkable number of problems in NLP. The course teaches different methods which can be applied to a wide range of tasks like document classification, word sense disambiguation, part-of-speech tagging, named entity recognition, parsing, discourse analysis, and language modeling.

COURSE OUTCOMES

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

CO1 : To understand basics of linguistics, probability and statistics
CO2 : To study statistical approaches to NLP and understand sequence labeling
CO3 : To outline different parsing techniques associated with NLP
CO4 : To explore semantics of words and semantic role labeling of sentences
CO5 : To understand discourse analysis, and opinion mining

UNIT 1: INTRODUCTION **8Hrs**

Overview: Origins and challenges of NLP Language and Grammar-Processing Indian Languages- NLP Applications Information Retrieval. Language Modeling: Various Grammar-based Language Models-Statistical Language Model. Natural Language Processing – Components - Basics of Linguistics and Probability and Statistics – Words-Tokenization-Morphology-Finite State Automata.

UNIT 2 : N-GRAMS AND LANGUAGE MODELS **8Hrs**

Smoothing -Text classification- Naïve Bayes classifier Evaluation - Vector Semantics – TF-IDF - Word2Vec- Evaluating Vector Models -Sequence Labeling – Part of Speech – Part of Speech Tagging -Named Entities –Named Entity Tagging.

UNIT 3 : CONTEXTUAL EMBEDDING **8Hrs**

Constituency –Context Free Grammar –Lexicalized Grammars- CKY Parsing – Earley's algorithm Evaluating Parsers -Partial Parsing – Dependency Relations- Dependency Parsing - Transition Based - Graph Based

UNIT 4 : COMPUTATIONAL SEMANTICS **8Hrs**

Word Senses and WordNet – BERT, Word Sense Disambiguation – Semantic Role Labeling – Proposition Bank- FrameNet- Selectional Restrictions - Information Extraction - Template Filling

UNIT 5: DISCOURSE ANALYSIS **8Hrs**

Discourse Coherence – Discourse Structure Parsing – Centering and Entity Based Coherence – Question Answering –Factoid Question Answering – Classical QA Models, Sentiment analysis – Polarity detection, RNN, Limitation of RNN for NLP problmes, Long short term memory networks.

SELF STUDY **5Hrs**

Text Books:

1. Tanveer Siddiqui, U.S. Tiwary, “Natural Language Processing and Information Retrieval”, Oxford University Press, 2008.
2. Anne Kao and Stephen R. Poteet (Eds), “Natural Language Processing and Text Mining”, Springer-Verlag London Limited 2007.

Reference Books:

1. Daniel Jurafsky and James H.Martin, “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition” (Prentice Hall Series in Artificial Intelligence), 2020
2. Jacob Eisenstein. “Natural Language Processing “, MIT Press, 2019
3. Samuel Burns “Natural Language Processing: A Quick Introduction to NLP with Python and NLTK, 2019
4. Christopher Manning, “Foundations of Statistical Natural Language Processing”, MIT Press, 2009.
5. Nitin Indurkhya,Fred J. Damerau, “Handbook of Natural Language Processing”, Second edition, Chapman & Hall/CRC: Machine Learning & Pattern Recognition, Hardcover,2010
6. Deepti Chopra, Nisheeth Joshi, “Mastering Natural Language Processing with Python”, Packt Publishing Limited, 2016
7. Mohamed Zakaria Kurdi “Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax (Cognitive Science)”, ISTE Ltd., 2016.

Semester	SECOND
Paper Code	BDADE 2724
Paper Title	DIGITAL IMAGE 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. To become familiar with digital image fundamentals

2. To get exposed to simple image enhancement techniques in Spatial and Frequency domain.
3. To study the image segmentation and representation techniques.
4. To become familiar with image compression and recognition methods

COURSE OUTCOMES:

CO1: Know and understand the basics and fundamentals of digital image processing, such as digitization, sampling, quantization, and 2D-transforms.

CO2: Operate on images using the techniques of smoothing, sharpening and enhancement.

CO3: Learn the basics of segmentation, features extraction and compression

CO4: Learn applications of image processing.

UNIT 1: FUNDAMENTALS OF IMAGE PROCESSING 10 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 2: IMAGE ENHANCEMENT 10 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 3: IMAGE COMPRESSION & SEGMENTATION 10 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.

UNIT 4: APPLICATIONS OF IMAGE PROCESSING 10 Hrs.

Image Classification – Image Recognition – Image Understanding – Video Motion Analysis – Image Fusion – Steganography – Digital Compositing – Mosaics – Colour Image Processing

SELF STUDY 5 Hrs.

REFERENCE BOOKS

1. Rafael C.Gonzalez and Richard E.Woods, “Digital Image Processing” Second Edition, Pearson Education, 2003
2. Milan Sonka, Vaclav Hlavac and Roger Boyle, “Image Processing, Analysis and Machine Vision”, Second Edition, Thomson Learning, 2001
3. Anil K.Jain, “Fundamentals of Digital Image Processing”, PHI, 2006.

Semester	SECOND
Paper Code	BDADE 2821
Paper Title	RESEARCH METHODOLOGY
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

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 1: CONCEPT OF RESEARCH **10 hours**

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 2: RESEARCH DESIGN AND SAMPLE SURVEYS **12 hours**

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 3: DATA COLLECTION AND DATA PROCESSING **12 hours**

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 4: REPORT WRITING & RESEARCH ETHICS, IPR AND PUBLISHING

11 hours

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

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 Machine Learning I LAB

1. Linear Regression
2. Logistic Regression
3. Neural Networks(MLP)
4. Support Vector Machines
5. Unsupervised Learning(K-means)
6. Dimensionality Reduction(PCS, TSNE)

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)

Code number and Title of the paper : BD2P5 NLP LAB

PRACTICAL EXERCISES:

1. Download nltk packages.
   Use it to print the tokens in a document and the sentences from it.
2. Include custom stop words and remove them and all stop words from a given document using nltk package.
3. Implement a stemmer and a lemmatizer program.
4. Implement a simple Part-of-Speech Tagger
5. Write a program to calculate TFIDF of documents and find the cosine similarity between any two documents.
6. Use nltk to implement a dependency parser.
7. Implement a semantic language processor that uses WordNet for semantic tagging.
8. Project must use NLP concepts and apply them to some data.

a. Collecting textual data
b. Cleaning the input text data
c. Normalizing the data
d. Feature engineering
e. Apply machine learning techniques to build models
f. Validating the model built
g. Deploying the model and making prediction on new data.

Code number and Title of the paper : BD2P6 DIGITAL IMAGE PROCESSING LAB

List of Experiments(Any 10 experiments)

1. Display of Gray scale Images.
2. Histogram Equalization.
3. Design of Non-linear Filtering
4. Determination of Edge detection using Operators
5. Contrast stretching of a low contrast image, Histogram, and Histogram Equalization
6. 2-D DFT and DCT.
7. Filtering in frequency domain.
8. Conversion between colour spaces.
9. Display of FFT(1-D & 2-D) of an image
10. DWT of images.
11. Implementation of Image Smoothening Filters(Mean and Median filtering of an Image)
12. Implementation of image sharpening filters and Edge Detection using Gradient Filters
13. Image Compression by DCT,DPCM, HUFFMAN coding
14. Implementation of image restoring techniques
15. Implementation of Image Intensity slicing technique for image enhancement
16. Segmentation using watershed transform.

SEMESTER III

Semester	THIRD
Paper Code	BDA 3124
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

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	**7 Hrs.**
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 3224
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 3324
Paper Title	MACHINE LEARNING II
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 advanced concepts and algorithms of deep learning, as well as recent research topics.

COURSE OBJECTIVES:

- To have a thorough understanding 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:

1. Develop and Train Deep Neural Networks
2. Develop a CNN, R-CNN, Fast R-CNN for detection and recognition.
3. Work with NLP and Word Embedding's.
4. 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.

UNIT II: CONVOLUTIONAL NEURAL NETWORKS 10 Hrs.

About 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 III: NLP & GENETIC ALGORITHMS 10 Hrs.

Text Classification: Processing raw text, Categorizing and Tagging words, from text to tokens, Text Classification, Applications: Summarization, Question Answering

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

UNIT IV: GRAPHICAL MODELS 10 Hrs.

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 3424
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	BDADE 3524
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:

Upon successful completion of this course student should be able to

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

Semester	SECOND
Paper Code	BDADE 3624
Paper Title	DIGITAL SIGNAL 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 signal Processing and its applications.

COURSE OBJECTIVES:

To impart knowledge about the following topics:

1. Signals and systems & Transformation techniques
2. Discrete time systems.
3. Their mathematical representation. & Programmability digital signal processor
4. Their computation. & quantization effects

COURSE OUTCOMES:

CO1: Ability to acquire knowledge on Signals and systems & their mathematical representation.

CO2: Ability to understand and analyze the discrete time systems.

CO3: Ability to analyze the transformation techniques & their computation

CO4: Ability to acquire knowledge on programmability digital signal processor & quantization effects.

UNIT 1: INTRODUCTION **10 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 TIME SYSTEM ANALYSIS **10 Hrs.**

Z-transform and its properties, inverse z-transforms; difference equation – Solution by z transform, application to discrete systems – Stability analysis, frequency response – Convolution – Discrete Time Fourier transform

UNIT 3: DISCRETE FOURIER TRANSFORM & COMPUTATION **10 Hrs.**

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

UNIT 4: DIGITAL SIGNAL PROCESSORS **10 Hrs.**

Introduction – Architecture – Features – Addressing Formats – Functional modes – Introduction to Commercial DS Processors.

SELF STUDY **5 Hrs.**

REFERENCE BOOKS

4. J.G. Proakis and D.G. Manolakis, 'Digital Signal Processing Principles, Algorithms and Applications', Pearson Education, New Delhi, PHI. 2003.
5. S.K. Mitra, 'Digital Signal Processing – A Computer Based Approach', McGraw Hill Edu, 2013.
6. Lonnie C.Ludeman ,"Fundamentals of Digital Signal Processing",Wiley,2013

LABORATORY

BD3P1: MODELING IN OPERATION MANAGEMENT LAB

Project

BD3P2: ENABLING TECHNOLOGIES FOR DATA SCIENCE II LAB

List of Experiments

Spark

Scala

Mahout

BD3P3: MACHINE LEARNING II Lab

Project

BDDE3P4: INTRODUCTION TO ECONOMETRICS AND FINANCE LAB

Panel data

GMM

Models(ARCH/GARCH/SV)

BDA3P5: Research Paper oriented LAB

Research Paper/ Research Project