ST.JOSEPH'SCOLLEGE (AUTONOMOUS)

BENGALURU-27

Re-accredited with **'A++' GRADE with 3.79/4 CGPA** by
NAAC Recognized by UGC as College of Excellence

ST. JOSEPH'S INSTITUTE OF INFORMATION TECHNOLOGY

DEPARTMENT OF ADVANCED COMPUTING

SYLLABUS FOR POSTGRADUATE PROGRAMME

SUMMARY OF CREDITS IN MSC(BIG DATA ANALYTICS)

Department of Advance Computing (PG)
2021-2024

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 1121	Basic Statistical Methods	45	03	03	70-2.5Hrs
Theory	BDA 1221	Probability & Stochastic Process	45	03	03	70-2.5Hrs
Theory	BDA 1321	Linear Algebra & Linear Programming	45	03	03	70-2.5Hrs
Theory	BDA 1421	Computing for Data Science	45	03	03	70-2.5Hrs
Theory	BDA 1521	Database Management System	45	03	03	70-2.5Hrs
Theory	BDA 1621	Python Programming	45	03	03	70-2.5Hrs
Practical	BD1P1	Basic Statistical Methods Lab	30	02	01	70-2.5Hrs
Practical	BD1P2	Probability & Stochastic Process Lab	30	02	01	70-2.5Hrs
Practical	BD1P3	Linear Algebra & Linear Programming Lab	30	02	01	70-2.5Hrs
Practical	BD1P4	Computing for Data Science Lab	30	02	01	70-2.5Hrs
Practical	BD1P5	Database Management System Lab	30	02	01	70-2.5Hrs
Practical	BD1P6	Python Lab	30	02	01	70-2.5Hrs
Total Number of credits:			**24**			

Semester 2	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of teaching hrs /week**	**Number of credits**	**Continuous Internal Assessment (CIA) Marks**
Theory	BDA 2121	Foundation of data Science (Programming for Big Data)	45	03	03	70-2.5Hrs
Theory	BDA 2221	Advance Statistical Method	45	03	03	70-2.5Hrs
Theory	BDA 2321	Machine Learning I	45	03	03	70-2.5Hrs
Theory	BDA 2421	Enabling Technologies for Data Science I	45	03	03	70-2.5Hrs
Theory	BDA 2521	Value thinking	45	03	03	70-2.5Hrs
Theory (DE)	BDADE 2621	Multivariate Statistics	45	03	03	70-2.5Hrs
Theory (DE)	Swayam	Computer Vision			03	
Theory (DE)	Swayam	Natural Language Processing			03	
Theory (DE)	BDADE 2721	Digital Image Processing	45	03	03	70-2.5Hrs
Theory (DE)	BDADE 2821	Digital Signal Processing	45	03	03	70-2.5Hrs
Practical	BD2P1	Foundation of data Science (Programming for Big Data) Lab	30	02	01	70-2.5Hrs
Practical	BD2P2	Advance Statistical Method Lab	30	02	01	70-2.5Hrs
Practical	BD2P3	Machine Learning I Lab	30	02	01	70-2.5Hrs
Practical	BD2P4	Enabling Technologies for Data Science I Lab	30	02	01	70-2.5Hrs
Practical	BDDE2	Multivariate Statistics	30	02	01	70-2.5Hrs

(DE)	P5	Lab				
Total Number of credits:			**26**			
Semester 3	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of teaching hrs /week**	**Number of credits**	**Continuous Internal Assessment (CIA) Marks**
Theory	BDA3121	Modeling in Operations Management	45	03	03	70-2.5Hrs
Theory	BDA3221	Enabling Technologies for Data Science II	45	03	03	70-2.5Hrs
Theory	BDA 3321	Machine Learning II	45	03	03	70-2.5Hrs
Theory	BDA 3421	Data Analytics on Cloud	45	03	03	70-2.5Hrs
Theory (DE)	BDADE 3521	Introduction to Econometrics and Finance	45	03	03	70-2.5Hrs
Theory (DE)	BDADE 3621	Bioinformatics	45	03	03	70-2.5Hrs
Practical	BD3P1	Modeling in Operation Management Lab	30	02	01	70-2.5Hrs
Practical	BD3P2	Enabling Technologies for Data Science II Lab	30	02	01	70-2.5Hrs
Practical	BD3P3	Machine Learning II Lab	30	02	01	70-2.5Hrs
Practical	BD3P4	Cloud Platform(s)	30	02	01	70-2.5Hrs
Practical	BDDE3 P5	Introduction to Econometrics and Finance Lab	30	02	01	70-2.5Hrs
Practical	BDDE3 P6	Bioinformatics Lab	30	02	01	70-2.5Hrs

	BD3P7	Research Oriented Project/ Paper	30	02	01	70-2.5Hrs
Total Number of credits:			**25**			
Semester 4	**Code Number**	**Title**	**No. of Hours of Instructions**	**Number of teaching hrs /week**	**Number of credits**	**Continuous Internal Assessment (CIA) Marks**
Practical	BDA4IN21	Internship based project	750		25	
		IGNITORS/ OUTREACH			04	
Total Number of credits:			**29**			

CORE COURSES (CC)	
Course Title	Code Number
Basic Statistical 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
Modelling 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	BDDE2P5
Modelling 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	BDDE3P1
Bioinformatics Lab	BDDE3P4

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	BDA1121
Paper Title	BASIC STATISTICAL 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 basic concepts of statistical methods and develop analytical ability to solve real-world problems using these methodologies.

COURSE OUTCOMES:

CO1: Understand the concept of data collection and analysis.

CO2: Design effective data visualizations in order to provide new insights and communicate information to the viewer.

CO3: Knowledge of Statistical techniques and its scope and importance

CO4: Discuss basic ideas of linear regression and correlation and their applications.

UNIT 1: DATA COLLECTION **13 Hrs.**

Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with date collection systems with examples from business, cleaning and treatment of missing data, Sampling techniques.

UNIT 2: DATA VISUALIZATION **7 Hrs.**

Principles of data visualization and different methods of presenting data in business analytics

UNIT 3: BASIC STATISTICS **10 Hrs.**

Frequency table, histogram, measures of location, measures of spread, skewness, curtosis, percentiles, box plot, relative frequency distribution as a statistics model

UNIT 4: CORRELATION AND REGRESSION **10 Hrs.**

Covariance, Correlation coefficient, properties of Correlation coefficient, Rank correlation, linear regression (two variables), Multiple correlation and partial correlation.

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

BLUE PRINT

Code number: BDA1120

Title of the paper: BASIC STATISTICAL METHODS

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	13	30
Unit II	7	20
Unit III	10	30
Unit IV	10	30
Self-Study	5	
TOTAL	**45**	**110**

Maximum marks for the paper (Excluding bonus question)= 70

Semester	FIRST
Paper Code	BDA1221
Paper Title	PROBABILITY & STOCHASTIC PROCESS
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVES:

The main objective of this course is to provide students with the foundations of probabilistic, stochastic process and time series analysis used in varied applications like disease modeling, climate prediction and computer networks etc.

COURSE OUTCOMES:

CO1: Understand the fundamental concepts of Probability that help to measure uncertainty

CO2: Understand the fundamental concepts of random variables and probability distribution

CO3: Understand the core concept of Test of Hypotheses

CO4: Understand the core concept of stochastic process

UNIT 1 BASIC PROBABILITY 8 Hrs.

Concepts of experiments, Outcomes, Sample space, Events, Combinatorial probability, Birthday paradox, Principle of inclusion & exclusion, Conditional probability, Independence, Bayes Theorem.

UNIT 2 PROBABILITY DISTRIBUTION: 12 Hrs.

Random Variables : discrete and continuous probability models, some probability distributions : Binomial, Poisson, Geometric, Hyper geometric, Gamma distribution, Beta Distribution, Normal, exponential, Chi-square, t-distribution, F-distribution, expectation, Variance and other properties of the distribution, Central Limit Theorem.

UNIT 3 TEST OF HYPOTHESES 13 Hrs.

Two types of errors, test statistic, parametric and non-parametric tests for equality of means and variances

UNIT4 STOCHASTIC PROCESS **7 Hrs.**

Markov Chains, Classification of states, Stationery distribution, Examples/case studies of stochastic process, Poisson process, illustrations and applications.

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

BLUE PRINT

Code number: BDA1221

Title of the paper: PROBABILITY & STOCHASTIC PROCESS

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	8	30
Unit II	12	30
Unit III	13	30
Unit IV	7	20
Self-Study	5	
TOTAL	**60**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	FIRST
Paper Code	BDA1321
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.

BLUE PRINT

Code number: BDA1321

Title of the paper: LINEAR ALGEBRA & LINEAR PROGRAMMING

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	12	30
Unit II	10	30
Unit III	10	30
Unit IV	8	20
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	FIRST
Paper Code	BDA1421
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

BLUEPRINT

Code number: BDA1421

Title of the paper: COMPUTING FOR DATA SCIENCES I

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	6	20
Unit II	13	40

Unit III	13	35
Unit IV	8	15
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	First
Paper Code	BDA1521
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.

BLUE PRINT

Code number: BDA1521

Title of the paper: DATABASE MANAGEMENT SYSTEM

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	10	20
Unit II	10	30
Unit III	10	30

Unit IV	10	30
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	FIRST
Paper Code	BDA1621
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.

BLUE PRINT

Code number: BDA1621

Title of the paper: Python Programming

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	8	20
Unit II	8	20
Unit III	8	30
Unit IV	8	20
Unit V	8	20
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

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

List of Programs -

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

Code number and Title of the paper : BD1P2 PROBABILITY & STOCHASTIC PROCESS LAB

1. Project using R shiny

Code number and Title of the paper : BD1P3 LINEAR ALGEBRA & LINEAR PROGRAMMING LAB

Different visualisation modules of Python

Code number and Title of the paper : BD1P4 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 : BD1P5 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 : BD1P6 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 2121
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.

BLUE PRINT

Code number: BDA 2121

Title of the paper: Foundation of Data Science

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	10	20
Unit II	10	30
Unit III	10	30
Unit IV	10	30
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	SECOND
Paper Code	BDA 2221
Paper Title	ADVANCE STATISTICAL METHOD
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 advance concepts of statistical methods and develop analytical ability to solve real-world problems using these methodologies.

COURSE OUTCOMES:

CO1: Understand the concept of estimation

CO2: Understand the importance of hypothesis testing

CO3: Understand the concept of ANOVA for data analysis

CO4: Use regression for building models

.

UNIT 1: ESTIMATION **16 Hrs.**

Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates, Expectation maximization algorithm, bootstrap algorithm.

UNIT 2: LINEAR MODEL **16 Hrs.**

Gauss Markov Model, least square estimators, Multiple linear regressions, forward, backward & stepwise regression, Logistic Regression, Polynomial regression, Analysis of variance.

UNIT 3: INTRODUCTION TO TIME SERIES **8 Hrs.**

Components of time series, Smoothing auto correlation, stationarity, concepts of AR, MA, ARMA & ARIMA models with illustrations

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

BLUE PRINT

Code number: BDA 2221

Title of the paper: Advance Statistical Method

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	16	45
Unit II	16	45
Unit III	8	20
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	SECOND
Paper Code	BDA2321

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.

BLUE PRINT

Code number: BDA2321

Title of the paper: Machine Learning I

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	10	20
Unit II	10	30
Unit III	10	30
Unit IV	10	30
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	SECOND
Paper Code	BDA 2421
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

BLUE PRINT

Code number: BDA 2421

Title of the paper: Enabling Technologies for Data Science I

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	12	30
Unit II	12	30
Unit III	6	20
Unit IV	5	20
Unit V	5	10
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	SECOND
Paper Code	BDA 2521
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 2621
Paper Title	MULTIVARIATE STATISTICS
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 concepts of multivariate statistical methods and develop analytical ability to solve real-world problems using these methodologies in data analytics field.

COURSE OUTCOMES:

CO1:Understand the concept of Multivariate data and its application

CO2:Understand the importance of Principal Component Analysis

CO3:Gain insight knowledge of the concept of Classification of data for data analysis

CO4:Understand the concept of Factor analysis

CO5:Acquire knowledge of the concepts of Clustering

UNIT 1: MULTIVARIATE DATA **15 Hrs.**

Representation of multivariate data, bivariate and multivariate distributions, multinomial distribution, multivariate normal distribution, sample mean & sample dispersion matrix, concepts of location & depth in multivariate data.

UNIT 2: PRINCIPAL COMPONENT ANALYSIS **6 Hrs.**

Population Principal Components, Summarizing Sample Variation by Principal Components, Graphing the Principal Components, Monitoring Quality with Principal Component

UNIT 3: CLASSIFICATION **8 Hrs.**

Separation and classification for 2 population, Classification with 2 multivariate normal distributions, Evaluating classification functions, Classification with several population, Loogistic Regression and Classification

UNIT 4: FACTOR ANALYSIS **5 Hrs.**

Orthogonal Factor model, Methods of Estimation, Factor rotation, Factor stores, Perspectives and Strategy for Factor Analysis

UNIT 5: CLUSTERING **6 Hrs.**

Similarity Measures, Hierarchical Clustering Methods, Non hierarchical clustering methods, Clustering based on Statistical models, Multidimensional scaling, Correspondence Analysis

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS :

1. Applied Multivariate Statistical Analysis : Richard A. Johnson and Dean W. Wichern, Prentice Hall, 2002

BLUE PRINT

Code number: BDADE 2621

Title of the paper: Multivariate Statistics

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	15	30
Unit II	6	20
Unit III	8	20
Unit IV	5	20
Unit V	6	20
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	SECOND
Paper Code	BDADE 2721
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

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

BLUE PRINT

Code number: BDADE 2821

Title of the paper: Digital Signal Processing

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	10	30
Unit II	10	30
Unit III	10	30
Unit IV	10	20
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	SECOND
Paper Code	BDADE 2721
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

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.

BLUE PRINT

Code number: BDADE 2721

Title of the paper: Digital Image Processing

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	10	30
Unit II	10	30
Unit III	10	30
Unit IV	10	20
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

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 Advance Statistics 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 – 1.
9. Applied Regression Analysis – 2.

10. Applied Regression Analysis – 3.
11. Logistic Regression.

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

Code number and Title of the paper : BD2P5 Multivariate statistics LAB

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

SEMESTER III

Semester	THIRD
Paper Code	BDA 3121
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

BLUE PRINT

Code number: BDA 3120

Title of the paper: Modeling in Operations Management

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	5	10
Unit II	7	20
Unit III	7	20
Unit IV	7	20
Unit V	7	20
Unit VI	7	20
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	THIRD
Paper Code	BDA 3221
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.

BLUE PRINT

Code number: BDA 3221

Title of the paper: Enabling Technologies for Data Science II

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	5	10
Unit II	10	30
Unit III	10	30
Unit IV	10	30
Unit V	5	10
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	THIRD
Paper Code	BDA 3321
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 OBJECTIVES:

- To have a thorough understanding of the Supervised and Unsupervised learning techniques
- To study the various probability based learning techniques
- To understand graphical models of machine learning algorithms
- To have basic understanding of genetic algorithms

COURSE OUTCOMES:

CO1: Suggest supervised, unsupervised or semi-supervised learning algorithms for any given problem.

CO2:Understand the concept of genetic algorithms and combining models.

CO3:Modify existing machine learning algorithms to improve classification efficiency

CO4:Design systems that use the appropriate graph models of machine learning

UNIT 1: DEEP LEARNING MODELS **13 Hrs.**

Perceptron algorithms, , Multi-layer Perceptron, Neural Networks,. Pattern recognition models.

UNIT 2: CLASSIFIERS & CLUSTERING **7 Hrs.**

Probabilistic Classifiers: Generative and Conditional classifiers. Hyperplane classifiers: Loss functions, Stochastic gradient algorithms,. Clustering:Performance criteria, K-means clustering,

UNIT 3: 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.

Text Generation and GAs

UNIT 4: GRAPHICAL MODELS **10 Hrs.**

Markov Chain Monte Carlo Methods – Sampling – Proposal Distribution – Markov Chain Monte Carlo – Graphical Models – Bayesian Networks – Markov Random Fields – Hidden Markov Models – Tracking Methods

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.

BLUE PRINT

Code number: BDA3321

Title of the paper: Machine Learning II

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	13	30
Unit II	7	20
Unit III	10	30
Unit IV	10	30
Self-Study	5	
TOTAL		**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	THIRD
Paper Code	BDADE 3521
Paper Title	ECONOMETRICS
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

BLUE PRINT

Code number: BDADE 3521

Title of the paper: Econometrics

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	14	40
Unit II	14	40
Unit III	5	20
Unit IV	3	10
Unit V	4	10
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	THIRD
Paper Code	BDA 3421
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.

BLUE PRINT

CODE NUMBER: BDA 3421

TITLE OF THE PAPER: Data analytics on cloud

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	10	20
Unit II	10	30
Unit III	10	30
Unit IV	10	30
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

Semester	THIRD
Paper Code	BDADE 3621
Paper Title	BIOINFORMATICS
Number of teaching hrs per week	3 Hrs
Total number of teaching hrs per semester	45
Number of credits	3

COURSE OBJECTIVE:

This aim of the course is to provide an overview of the functioning of biological systems along with the advanced knowledge on sequence analysis, biological data collection, interpretation, and analysis.

COURSE OUTCOMES:

Upon successful completion of this course student will gain

CO1: knowledge on functioning of biological systems.

CO2: Understand the advanced concepts of sequence analysis. In addition,

CO3: Capable of collection of various types of biological data and their interpretation and analysis.

CO4: Develop analytical ability to solve real-world problems using these methodologies.

UNIT 1: OVERVIEW ON CELL AMD MOLECULAR BIOLOGY **7 Hrs.**

UNIT 2: GENOMICS, TRANSCRIPTOMICS, AND PROTEOMICS **7 Hrs.**

UNIT 3: BIOLOGICAL DATA COLLECTION **7 Hrs.**

UNIT 4: SEQUENCE ALIGNMENT AND ALGORITHMS **9 Hrs.**

UNIT 5: ADVANCED METHODS OF SEQUENCE ANALYSIS **7 Hrs.**

UNIT 6: PROTEIN STRUCTURE PREDICTION **3 Hrs.**

SELF STUDY **5 Hrs.**

SUGGESTED BOOKS:

1. Introduction to Computational Molecular Biology : C. Setubal & J. Meidanis, PWS Publishing, Boston, 1997

BLUEPRINT

Code number: BDADE3621

Title of the paper: Bioinformatics

Chapter	Number of Hours	Total marks for which the questions are to be asked (including bonus questions
Unit I	7	20
Unit II	7	20

Unit III	7	20
Unit IV	9	20
Unit V	7	20
Unit VI	3	10
Self-Study	5	
TOTAL	**45**	**110**
Maximum marks for the paper (Excluding bonus question)= 70		

LABORATORY

BD3P1: Modelling 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

BD3P4: Cloud Platform Lab

Numpy

Keras

Pytorch

BDDE3P5: Introduction to Econometrics and Finance LAB

Panel data

GMM

Models(ARCH/GARCH/SV)

BDDE3P6: Bioinformatics LAB

Applications of Bioinformatics

BDA3P7: Research Paper oriented LAB

Research Paper

List of Experiments

1. Retrieving, analyzing and interpreting gene and protein related data.
2. Advanced analysis and visualization of biological data.
3. Sequence alignment and analysis.