SRI SARADA COLLEGE FOR WOMEN (AUTONOMOUS),

Reaccredited with 'B++' Grade by NAAC

Affiliated to Periyar University
Fairlands, SALEM- 636 016.

OUTCOME BASED SYLLABUS

PG & RESEARCH DEPARTMENT OF COMPUTER SCIENCE

M.Sc. DATA SCIENCE
(Self-Financing)

(For the students admitted in 2022-2023)

M. Sc. DATA SCIENCE
PROGRAMME OUTCOMES

PO1   To provide knowledge of theoretical data science which can be employed in research and complex business situations

PO2   To identify, formulate and solve problems to meet a given set of computing requirements in the context of the programme discipline.

PO3   To design, implement, and evaluate a computer-based system, process, component or programme using ethical codes and procedures.

PO4   To keep the students abreast of the latest development in the field of data science.

PO5   To help students in creating programmes to process information with team working skills.

SRI SARADA COLLEGE FOR WOMEN (AUTONOMOUS), SALEM-16
PG & RESEARCH DEPARTMENT OF COMPUTER SCIENCE
M.Sc. DATA SCIENCE
PROGRAMME STRUCTURE UNDER CBCS
(For the students admitted in 2022-2023)
Total Credits: 92 + Extra Credit (Maximum 16)

SEMESTER – I				
Course	**Course Title**	**Code**	**Hrs./ Week**	**Credits**
Core - I	Foundations of Data Science	22PDSC1	5	5
Core - II	Data Structures and Algorithms	22PDSC2	5	5
Core - III	Python for Data Analytics	22PDSC3	5	5
Elective - I	Statistics for Data Science / Business Analytics	22PDSEC1 / 22PDSEC1A	4	4
Core Practical - I	Data Structures Lab	22PDSQC1	5	4
Core Practical - II	Python Programming Lab	22PDSQC2	6	5
Total			**30**	**28**
Extra Skills	* *Value Education* * *Physical Fitness Practice* * *Life Skills Promotion* * *Productive Preparation for CBSE NET/SET/JRF - I (21PDSSC1)* *(Self-Study - 1 Extra Credit)*			

SEMESTER-II				
Course	**Course Title**	**Code**	**Hrs./ Week**	**Credits**
Core - IV	Data Visualization	22PDSC4	5	5
Core - V	Data Mining with R	22PDSC5	5	5
Core - VI	Mathematical Foundations for Data Science	22PDSC6	5	5
Elective - II	Text Analytics/ Web Analytics	22PDSEC2 / 22PDSEC2A	4	4
Core Practical - III	Visualization Lab	22PDSQC3	5	5
Core Practical - IV	R Programming Lab	221PDSQC4	4	4
	Human Rights	22PHRSC	2	2
Total			**30**	**30**
Extra Skills	* *Value Education - 1 Extra Credit* * *Physical Fitness Practice -1Extra Credit* * *Life Skills Promotion -1 Extra Credit* * *Productive Preparation for CBSE NET/SET/JRF - II (21PDSSC2)* *(Self-Study - 1 Extra Credit)* * *Extension Activity - 1Extra Credit* * *In-plant Training - During Semester Vacation - 1 Extra Credit*			

- *Extra credits are given for extra skills and courses qualified in MOOC/NPTEL*

SEMESTER-III				
Course	**Course Title**	**Code**	**Hrs./ Week**	**Credits**
Core - VII	Big Data Analytics	22PDSC7	5	5
Core - VIII	Machine Learning	22PDSC8	5	5
Elective - III	Cloud Computing / Distributed Databases	22PDSEC3/ 22PDSEC3A	4	4
Elective - IV	Internet of Things /Real Time Analytics	22PDSEC4 / 22PDSEC4A	4	4
Elective - V	Big Data Lab / Mini Project	22PDSEC5 / 22PDSEC5A	4	2
Core Practical - V	Machine Learning Lab	22PDSQC5	4	2
Extra - Disciplinary Course (EDC)	Applied Business Research / Management Accounting	22PCOEDC / 22PCOEDCA	4	4
	Total		**30**	**26**
Extra Skills	** Value Education* ** Life Skills Promotion* ** Physical Fitness Practice* ** Productive Preparation for CBSE NET/SET/JRF-III (21PDSSC3)* *(Self-Study -1 Extra Credit)*			

* Preparation for the Project - 5 Hours per Week (Out of College Hours)

SEMESTER-IV				
Course	**Course Title**	**Code**	**Hrs. /Week**	**Credits**
	External Project	22PDSPC	30	8
Extra Skills	** Value Education - 1 Extra Credit* ** Productive Preparation for CBSE NET/SET/JRF - IV (21PDSSC4)* *(Self-Study - 1 Extra Credit)* ** Extension Activity - 1 Extra Credit* ** Technical Aptitude for CBSE NET/SET/JRF : Paper - I - 1 Extra Credit*			
	Total		30	8
	Grand Total		**120**	**92**

- *Extra credits are given for extra skills and courses qualified in MOOC/NPTEL*

Extra-Disciplinary Courses offered by Dept. of Computer Science
1. Database Management Systems
2. Image Editing and Animation

Programme Title **:** **M.Sc. Data Science**

Course Title **:** **FOUNDATIONS OF DATASCIENCE**

Course Code **:** **22PDSC1** **Hours/Week:5**

Semester **:** **I** **Credit:5**

COURSE OBJECTIVES:

1. Introduce basis of data evolution and understanding of data
2. Explain the basic concepts of data science
3. Detail out the strategies related to big data
4. Teach about exploratory data analysis.
5. Learn about the applications of machine learning

SYLLABUS

Unit – I: Introduction (Hours: 15)

A brief history of data science -The history of data gathering- Where is data science used-Data and data set -Perspectives on data-The crisp DM process.

Unit – II: Data Science Eco-system (Hours: 15)

Moving the algorithms to the data- The Traditional Database or the Modern Traditional Database- Big Data Infrastructure- The Hybrid Database World- Data Preparation and Integration- Creating the Analytics Base Table.

Unit – III: Data science in a big data world (Hours: 15)

Benefits and uses of data science and big data-Facets of Data-The data science process-The big data ecosystem and data science.

Unit - IV: The Data Science process (Hours: 15)

Overview of the data science process-Defining goals-Retrieving data-Cleansing integrating and transforming data-Exploratory data analysis-Build the models-Presenting findings.

Unit - V: Machine Learning (Hours: 15)

Applications for machine learning in data science-Where machine learning is used in the data science process-Python tools used in machine learning-The modelling process-Types of machine learning.

BOOKS FOR STUDY:

1. John.D.Kelleher and Brenden Tierney, "Data Science", The MIT Press Essential Knowledge Series,2018.(Unit 1to 2)

2. Davy Cielen Arno D. B. Meysman Mohamed Ali ,"Introducing Data Science Big Data, Machine Learning, And More, Using PythonTools",Manning Shelter Island;2016.(Unit 3 to 5)

BOOKS FOR REFERENCE:

1. Lillian Pierson,"Data Science for Dummies", John Wiley & Sons, 2015.

2. EMC Education Services,"Data Science and Big Data Anlaytics",Wiley Publishing,2015.

WEB RESOURCES :

- https://www.youtube.com/watch?v=KxryzSO1Fjs

- https://www.youtube.com/watch?v=dMpdoprDEDI

- https://www.youtube.com/watch?v=ZqjBPYX_2bk

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Outline** the concepts of data and data sets	**K2**
CO2	**Apply** the skill to cleansing and transforming of data to build the model	**K3**
CO3	**Analyse**d the extensive concepts of exploratory data analysis	**K4**
CO4	**Estimate** the performance of machine learning concepts	**K6**

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	M	S
CO2	S	S	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M

S - Strong, M - Medium, L - Low

Programme Title **:** **M.Sc. Data Science**

Course Title **:** **DATA STRUCTURES AND ALGORITHMS**

Course Code **:** **22PDSC2** **Hours/Week:5**

Semester **:** **I** **Credit:5**

COURSE OBJECTIVES:

1. Provide the fundamentals of data structures and algorithms
2. Elucidate the algorithmic problem solving methods
3. Explain the concept of OOPs Language
4. Accustom the concepts of dynamic programming

SYLLABUS

Unit - I: Data Structures and Algorithms (Hours: 10)

The Need for Data Structures- Abstract data types and data structures-Problems, Algorithms and Programs-Algorithm Analysis- Best, Worst, and Average Cases- A Faster Computer, or a Faster Algorithm- Asymptotic Analysis- Calculating the Running Time for a Program.

Unit - II: Fundamental Data Structures (Hours: 15)

Stacks - How stacks are used-Stack ADT-Applications-Queues- How queues are used-Queue ADT-Tree-Binary tree types-Binary tree traversal-Graph-Applications of graph-Graph Representation.

Unit - III: Object Oriented Language (Hours: 15)

Object Oriented language fundamentals- programming basics- Conditional statements-Structures- Functions - Objects and Classes- Constructors- Overloading. Inheritance-Hierarchy - Derived class- Access specification - Polymorphism- virtual functions- virtual class- Files - Exception Handling.

Unit - IV: Design of Algorithms (Hours: 18)

Divide and Conquer: General Method, Binary Search, Merge sort, Quick sort. Greedy Method: Knapsack problem, Job sequencing with deadlines.

Unit - V: Dynamic Programming (Hours: 17)

Computing a Binomial Coefficient- Marshall's and Floyd algorithm- Optimal Binary Search Trees- Knapsack Problem and Memory functions. Greedy Technique-Prim's algorithm-Kruskal's Algorithm-Dijkstra's Algorithm.

BOOKS FOR STUDY:

1. Clifford A. Shaffer,"Data Structures and Algorithm Analysis", Edition 3.2 (C++ Version) ",Virginia Tech Blacksburg ,2013.
2. NarasimhaKarumanchi,"Data Structures and Algorithms Made Easy Edition 5",IIT Bombay,2017.
3. BjarneStroustrup, "The C++ Programming Language", Addison Wesley, 7th Edition, ISBN: 0321563840, 2017.
4. Sahni,"Data Structures,Algorithms, And Applications in C++", Second Edition,2010

BOOKS FOR REFERENCE:

1. Mark Allen Weiss,"Data Structures and Algorithm Analysis in C++";Pearson Education,2012.
2. Alfred V. Aho, John E. Hopcroft and Jeffrey D. Ullman, "Data Structures and Algorithms", Pearson Education, Reprint 2006.

WEB RESOURCES :

- https://www.youtube.com/watch?v=zg9ih6SVACc
- https://www.youtube.com/watch?v=zWg7U0OEAoE
- https://nptel.ac.in/courses/106/106/106106127/

COURSE OUTCOMES (CO) : **On Completion of course the Students will be able to**

CO Number	CO Statement	Knowledge Level
CO1	**Summarize** the concepts of ADT and algorithms	K2
CO2	Using data structures and algorithms to **solve** social connect problems	K3
CO3	**Analyse**s the types of algorithms	K4
CO4	**Assess** efficiency trade-off's among different data structure implementations	K5
CO5	**Estimate** various types of searching, sorting and graph traversal algorithms.	K6

Mapping of COs with POs :

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M
CO5	S	S	M	S	S

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title   :   PYTHON FOR DATA ANALYTICS

Course Code   :   22PDSC3   **Hours/Week:5**

Semester   :   I   **Credit:5**

COURSE OBJECTIVES:
1. Deliver fundamental knowledge in Python
2. Familiarize basic concepts in Python Programming
3. Explain various working libraries like numpy, pandas
4. Identify the key issues in Clustering, Validation and Optimization.

SYLLABUS

Unit - I: (Hours: 12)

Structure of Python Program-Underlying mechanism of Module Execution-Branching and Looping-Problem Solving Using Branches and Loops-Functions - Lists and Mutability-Problem Solving Using Lists and Functions.

Unit - II: (Hours: 15)

Understanding python's role in data science- Introducing python capabilities - working with numbers and logic - creating and using strings - storing data using sets, lists, and tuples.

Unit - III:(Hours : 15) Scientific libraries- Basics of Numpy and Scipy- Computation on NumPy and Scipy - basic statistical tools- Pandas - Basics of Pandas - data cleaning and analysis, Matplotlib and Seaborn -Scikitlearn- build machine learning models.,

Unit IV: (Hours: 15)

Visualizing the data - choosing the right graph - creating advanced scatterplots - Basic of Matplotlib and Seaborn - Bar graph, Pie chart, Box plot, Histogram, Line charts - visualizing graphs - understanding the tools.

Unit V: (Hours : 18)

Wrangling data - Stretching python's capabilities-exploring data analysis-clustering - Preforming Cross - validation, Selection and Optimization.

BOOKS FOR STUDY:
1. Jake Vander Plas, "Python Data Science Handbook: Essential Tools for Working with Data", O'Reilly Media, 2016,
2. Joel Grus, "Data Science from Scratch First Principles with python", O'Reilly ,2015 .
3. Alberto Boschetti, Luca Massaron, " Python Data Science essentials", 2nd Edition -

packt publishing Ltd, 2016.

4. John Paul Mueller , Luca Massaron, " Python for DataScience for DUMMIES", A Wiley Brand, 2015 .

BOOKS FOR REFERENCE:

- Dr.R.Nageshwararao, " Core Python Programming", Dream etch Press -2nd Edition, 2016.

- Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, AbhijitDasgupta , "Practical Data Science Cookbook", Packt publishing, 2014.

MOOC :

- https://onlinecourses.nptel.ac.in/noc21_cs33/preview

Web Resources :

- https://dimensionless.in/understanding-different-components-roles-in-data-science/

- https://towardsdatascience.com/the-three-cores-of-data-science-d58af0d7361e

- https://www.tutorialspoint.com/python/

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Recall** the basic notions and concepts of Python	K1
CO2	**Interpret** the basic logics and application development.	K2
CO3	**Utilize** different scientific libraries to code GUI applications.	K3
CO4	**Examine** various Visualization methods and data analysis.	K4
CO5	**Determine** the usability and problem solving capability.	K5

Mapping of COs with POs :

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M
CO5	S	S	M	S	S

Programme Title	:	**M.Sc. Data Science**
Course Title	:	**ELECTIVE- I STATISTICS FOR DATA SCIENCE**
Course Code	:	**22PDSEC1** **Hours /Week:4**
Semester	:	**I** **Credit:4**

COURSE OBJECTIVES:

1. To understand quantitative and qualitative data
2. To present the data using curves
3. Familiarize different measures of dispersion.
4. Acquaint Knowledge of various correlation
5. Acquire the knowledge to use regression

SYLLABUS

Unit - I :(Hours: 10)

Introduction: Nature and scope of Statistics, limitations of statistics-Types of data: Concept of population and sample, primary and secondary data, quantitative and qualitative data, discrete and continuous data, cross-sectional and time series data. Diagrammatic and Graphical representation of data - Line diagram, bar diagram, pie diagram and sub-divided bar diagram, Frequency distribution and cumulative frequency distribution and their graphical representations, Frequency polygon, histogram, frequency curves.

Unit - II :(Hours: 10)

Measures of Location : Definition of location-Arithmetic mean, Median, Mode, Geometric mean and Harmonic mean for individual observations, discrete series and continuous series data and their properties.

Unit - III :(Hours :12)

Measures of Dispersion : Definition of dispersion-Range, Quartile deviation, Mean deviation, Standard deviation, combined standard deviation, co-efficient of variation for individual observations, discrete series and continuous series data and their properties.

Unit - IV: (Hours: 15)

Measures of Skewness- Definition of skewness-Karl Pearson's coefficient of skewness, Bowley's coefficient of skewness, Kelly's coefficient of skewness. Measure of Kurtosis-Definition of kurtosis-Co-efficient of skewness and Kurtosis based on moments.

Unit - V: (Hours: 13)

Bivariate data-scatter diagram, Correlation-KarlPearson's coefficient of correlation –Rank correlation-Spearman's rank correlation coefficient. Regression Analysis: Introduction- Uses of regression analysis- regression lines- regression equations of X on Y and Y on X-regression equation in terms of correlation table-Applications.

BOOKS FOR STUDY:

1. S.C.Gupta&V.K.Kapoor,"Fundamentals of Mathematical Statistics,",Sultan Chand & Sons.2002.

2. R.Wilcox, "Basic Statistics", Oxford UniversityPress,2009.

3. Goon A.M, Gupta M.K., Dasgupta B ,"Fundamentals of Statistics" Volume I, World press,1998.

BOOKS FOR REFERENCE:

1. Murray R Spiegel and Larry J Stephens: "Statistics", Schaum's Outline, Fourth edition, 2008.

2. R.S.N. Pillai, "Statistics", S.ChandPublishingCompanyPvt Ltd,1992.

WEB RESOURCES :

- https://www.youtube.com/watch?v=4LTeDHmKtHE

- https://www.youtube.com/watch?v=tcusIOfI_GM

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Recall** the various types of data.	K1
CO2	**Identify** the statistical methods for real time data.	K3
CO3	**Examine** simple and advanced statistical methods to process complex problems in the real world.	K4
CO4	**Predict** the changes of data according to the trends	K6

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M

S - Strong, M - Medium, L - Low

Programme Title	:	**M.Sc. Data Science**

Course Title	:	**ELECTIVE- I BUSINESS ANALYTICS**	
Course Code	:	**22PDSEC1A**	**Hours/Week:4**
Semester	:	**I**	**Credit:4**

COURSE OBJECTIVES:

1. Able to understand the challenges in health care industry and top health care adaptations
2. Gain knowledge in Banking, Finance and Insurance
3. Understand analytics tools in telecommunications
4. Analyze the consume behavior in retail industry

SYLLABUS

Unit - I: (Hours :10)

Healthcare analytics- Introduction - Potential contributions - Challenges of healthcare industry - current and future state of healthcare analytics- top healthcare analytics adaptations

Unit - II: (Hours :15)

Banking and Finance: Systems of Banking- Commercial Banking- New Financial Services: Insurance Services- Types of Insurance- Housing Finance.

Unit - III: (Hours :15)

Telecommunication: Introduction - End-User Needs and Demands- Telecom Business

Unit - IV: (Hours :10)

Retail analytics- Understanding the new consumer- Marketing in a consumer - driven era - Managing the brand to drive loyalty.

Unit - V:(Hours :10)

Case studies: Walmart, Netflix, Facebook, Uber, Amazon, Kaggle

BOOKS FOR STUDY:

1. Dwight McNeill "A Framework for Applying Analytics in Healthcare", Pearson Education.2013.

BOOKS FOR REFERENCE:

1. Gomez Clifford,"Banking and Finance Theory Law and practice", PHI Learning.2011.
2. Anders Olsson, "Understanding Changing Telecommunications" ,Wiley Publications,2004.

3. Jennifer LeClaire, Danielle Dahlstrom, Vivian Braun,"Business analytics in Retail for dummies", 2 nd IBM Limited edition.
4. Alistair Croll Lean analytics: "Use Data to Build a Better Startup faster", O Reilly Publishers 2013.
5. Bernard Marr, "Big Data in Practice- How 45 successful companies used big data".2016.

WEB RESOURCES :

- https://nptel.ac.in/courses/110/105/110105089/
- https://nptel.ac.in/courses/110/107/110107092/

COURSE OUTCOMES (CO) : **On Completion of course the Students will be able to:**

CO Number	CO Statement	Knowledge Level
CO1	**Infer** the challenges of the business world	K2
CO2	**Build** and enhance business intelligence capabilities by adapting the appropriate technology.	K3
CO3	**Analyze** legal and ethical principles applied to contexts and environments of data science and decision making.	K4
CO4	**Predict** consumer behaviour in marketing industry	K5
CO5	**Recommend** the right product to the right person through business analytics	K6

Mapping of COs with POs:

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M
CO5	S	M	S	M	S

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title : CORE PRACTICAL- I   DATA STRUCTURES LAB

Course Code : 22PDSQC1   **Hours/Week: 5**

Semester : I   **Credit:4**

COURSE OBJECTIVES:
1. Introduce the important problem types in algorithm concepts.
2. Familiarize the different types of data structure methodology.
3. Build programs using different types of techniques
4. To enhance different problem solving skills

SYLLABUS

1. Write a C++ program to implement
    a. Quick Sort   b. Binary search
2. Write a C++ program to implement the following
    a. Stack ADT   b. Queue ADT
3. Write C++ programs to implement the following using a singly linked list.
    a. Stack ADT   b. Queue ADT
4. Write a C++ program to perform the following operations:
    a. Insert an element into a binary search tree.
    b. Delete an element from a binary search tree.
    c. Search for a key element in a binary search tree.
5. Write C++ programs that use recursive functions to traverse the given binary tree in
    a. Preorder   b. in order and   c. post order.

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to:

CO Number	CO Statement	Knowledge Level
CO1	**Demonstrate** the various data representation techniques in the real world.	K2
CO2	**Construct** linear and non-linear data structures	K3
CO3	**Analyze** the various algorithms based on their time and space complexity	K4
CO4	**Decide** suitable data structure to solve various computing problems	K5
CO5	**Develop** real-time applications using suitable data structure.	K6

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	S	M	M	M
CO2	S	S	M	M	S
CO3	S	M	S	S	S
CO4	S	S	M	M	M
CO5	S	S	M	M	S

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title : CORE PRACTICAL - II PYTHON PROGRAMMING LAB

Course Code : 22PDSQC2 **Hours/Week: 6**

Semester : I **Credit:5**

COURSE OBJECTIVES:

1. Introduce Python programming concepts.
2. Familiarize the concepts numbers, logics, strings, lists.
3. Build programs using file handling functions.
4. Provide basic knowledge on working with Scientific libraries.
5. Analyze and evaluate different methods of clustering and classification.

SYLLABUS

1. Program to find a) GCD of two numbers.

   b) Maximum of a list of numbers.

2. Program to find removing all duplicate elements in the list by using Naive method, list comprehension method, set method.

3. Program using Statistical analysis for mean, median, mode, variance, standard deviation, correlation and covariance.

4. Write a program for Addition, Subtraction, Multiplication, Division, using functions and Control structures.

5. Write a program for File handling functions open, read, write, close, find length.

6. Write a program for string functions in python.

7. Write a program for Visualization using Matplotlib.

8. Write a program for Visualization using Seaborn.

9. Write a program for k-mean Clustering .

10. Write a program for Classification using Scikitlearn.

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Demonstrate** basic programming structure in Python programming	**K2**
CO2	**Develop** and acquire the programming skills of core python.	**K3**
CO3	**Inspect** OOP concepts to produce real-time applications.	**K4**
CO4	**Determine** few machine learning methods to handle various real time problems solving scenarios.	**K5**
CO5	**Construct** various tools of data transformation to perform data analysis.	**K6**

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M
CO5	S	S	M	S	S

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title   : DATA VISUALIZATION

Course Code   : 22PDSC4   **Hours/Week: 5**

Semester   : II   **Credit:5**

COURSE OBJECTIVES:

1. Introduce the basis techniques and theory used in visualization.
2. Conduct exploratory data analysis using visualization.
3. Use knowledge of perception and cognition to evaluate visualization design alternatives
4. Understand why visualization is an important part of data analysis
5. Understand the components involved in visualization design.

SYLLABUS

UNIT - I: CONTEXT OF DATA VISUALIZATION (Hours :15)

Definition, Methodology, Visualization design objectives. Key Factors- Purpose, visualization function and tone, visualization design options- Data representation, Data Presentation, Seven stages of data visualization, widgets, data visualization tools. Mapping - Time series - Connections and correlations - Scatter plot maps - Trees, Hierarchies and Recursion- Networks and Graphs, Infographics.

UNIT- II: INTRODUCTION TO D3 (Hours :15)

Fundamental Technology -Drawing with data- Scales- Axes- Updates, Transition and Motion-Interactivity - Layouts -Geomapping- Exporting- Data to create Visualization with SVG - SVG- Styling CSS -Shapes- SVG Properties- SVG Text - Drawing- Transformations-Building Chart with SVG (Scalable Vector Graphics) - Shaping Web Pages- Selections-Attributes- Chaining Methods-Data Joins - Sizing- scales- axes- Loading- Filtering-Interactive Charts- Buttons using Data Join- Transition using Key

UNIT- III: D3-BASED REUSABLE CHART LIBRARY (Hours :15)

Introduction to D3- Setup and Deployment- Generate Chart- Customize Chart- How to Use APIs- Customize Style- Building Real time and Live Updating animated graphs with C3.

UNIT- IV: TABLEAU INTRODUCTION (Hours: 15)

Environment Setup- Navigation- File & Data Types. DATA SOURCE: Custom Data View-Extracting Data- Fields Operations- Editing Meta Data- Data Joining- Data Blending. Worksheets

UNIT - V: TABLEAU CHART (Hours :15)

Bar Chart- Line Chart- Pie Chart- Scatter Plot- Bubble Chart - Gantt Chart- Histograms - Waterfall Charts. ADVANCED: Dashboard- Formatting - Forecasting- Trend Lines

BOOKS FOR STUDY:

1. Ben Fry, "Visualizing Data: Exploring and Explaining Data with the Processing Environment", O'Reilly, 1st Edition, 2008.

2. Scott Murray, "Interactive data visualization for the web: An Introduction to Designing with D3", O'Reilly, 2nd Edition, 2017

3. Joshua N. Milligan, "Learning Tableau 2019: Tools for Business Intelligence, data prep, and visual analytics", Packt Publishing Limited, 2019.

BOOKS FOR REFERENCE:

1. Ritchie S. King - Visual Storytelling with D3- An Introduction to Data Visualization with D3, Addison-Wesley-Data Analytic Series, ISBN 10:0321933176

2. Elijah Meeks , D3.js in Action, Second Edition: Data visualization with JavaScript, Publisher: Manning Publications, 2017 , ISBN:9781617294488

WEB RESOURCE

- https://www.tutorialspoint.com/tableau/index.htm
- https://www.dashingd3js.com/table-of-contents
- https://c3js.org/gettingstarted.html
- https://www.udacity.com-Data Visualization and D3.J

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Define** the deterministic feature of the data	**K1**
CO2	**Illustrate** the interactive visualizations for the Web	**K2**
CO3	**Build** the methods and algorithms used to map data to graphical depictions	**K3**
CO4	**Discover** different domains of data Interpretation	**K4**
CO5	**Predict** various domain specific applications using advanced data visualization tools.	**K5**

Mapping of COs with POs:

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M
CO5	S	S	M	S	S

S - Strong, M - Medium, L - Low

Programme Title	:	**M.Sc. Data Science**
Course Title	:	**DATA MINING WITH R**
Course Code	:	**22PDSC5**

Hours/Week: 5

Semester	:	**II**

Credit:5

COURSE OBJECTIVES:
1. Provide fundamental knowledge in Data Mining
2. Introduce Exploratory Data Analysis using R Coding
3. Gain knowledge about Visualization

SYLLABUS

Unit – I (Hours :15)

DATA MINING- Introduction - Kinds of Data- Data Mining Functionalities - Interestingness of Patterns- Classification of Data Mining Systems- Data Mining Task Primitives - Integration of a Data Mining System with a Data Warehouse- Issues -Data Preprocessing - Data Cleaning-Data Integration and Transformation - Data Reduction- Data Cube Aggregation - Attribute Subset Selection- Dimensionality Reduction- Numerosity Reduction- DataDiscretization - Concept Hierarchy Generation for Numerical Data- Concept Hierarchy Generation for Categorical Data.

Unit- II (Hours :15)

ASSOCIATION RULE MINING AND CLASSIFICATION - Mining Frequent Patterns, Associations and Correlations- Basic Concepts-: Frequent Itemset Mining Methods- The Apriori Algorithm- Mining Various Kinds of Association Rules- Classification and Prediction - Classification by Decision Tree Induction : Decision Tree Induction **-** Bayesian Classification : Bayes' Theorem - Naïve Bayesian Classification- Rule Based Classification : Using IF-THEN Rules for Classification - Rule Extraction from a Decision Tree - Classification by Backpropagation : A Multilayer Feed-Forward Neural Network - Defining a Network Topology- Backpropagation- Prediction : Linear Regression - Nonlinear Regression.

Unit- III (Hours :15)

CLUSTER ANALYSIS: Introduction - Categorization of Major clustering Methods- Partitioning Methods : K means - Hierarchical Methods : Agglomerative and Divisive Hierarchical Clustering - Density-Based Methods-DBSCAN- Data Mining Applications.

Unit- IV (**Hours :15**)

R INTRODUCTION: What is R-Downloading and Installing R-. Getting Data into R: First Step in R: Typing in Small Datasets- Concatenating Data with c Function- Combining Variables with the c, cbind, rbind Functions - Vector Function -Matrix - Ddata frame- List - Importing Excel Data- Accessing Data from other Statistical Packages- Accessing the Database. Functions - The Attach Function- Exporting Data - The Tapply Function- The Supply and Lapply Function- The Summary and Table Function. Importing Data- Csv, Excel, Table, Xml, Json , Databases Conditional- Control flow- Loops- A Function with Multiple Arguments

Unit- V (Hours :15)

DATA EXPLORATION AND VISUALIZATION - Descriptive Statistics In R - Decision trees - data distributions using box and scatter plot **-** outliers using plot - Plot the histogram, bar chart and pie chart on sample data. -text mining.

BOOKS FOR STUDY:
1. Data Science and Big Data Analytics, EMC Education Services, WileyPublishers,
2. Learning Predictive Analytics with R, Eric Mayor, PACKET Publishing,2015
3. Frank Ohlhorst, Big Data Analytics Turning Big Data Into Big Money, Wiley Publishers, 2012 (UNITSI,II)

BOOKS FOR REFERENCE:
1. Dr. Mark Gardener, Beginning R- The Statistical Programming Language., WileyIndia Private Limited, Edition 2014 (UNITS III,IV,V)

WEB RESOURCES
- https://analyticstraining.com/best-free-resources-on-r/
- https://www.dataschool.io/resources/

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Interpret** the typical data mining methods and techniques	K1&K2
CO2	**Make use of** a statistical operations over different data sets	K3
CO3	**Discuss** different phases and their emerging applications.	K4

CO4	**Evaluate** a wide range of emerging and newly-adopted methodologies and technologies to facilitate the knowledge discovery.	K5

Mapping of COs with POs :

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M

S - Strong, M - Medium, L - Low

Programme Title   :   M.Sc. Data Science

Course Title   :   MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE

Course Code   :   22PDSC6   **Hours/Week: 5**

Semester   :   II   **Credit:5**

COURSE OBJECTIVES:
1. Provide basic understanding of linear algebra.
2. Understand the basic concepts of matrices.
3. Explore the use of probability and statistics for data science
4. Evaluate the various optimization techniques to solve problems.

SYLLABUS

Unit I: (Hours: 12)

Basics of Data Science - Introduction - Typology of problems - Importance of linear algebra, statistics and optimization from a data science perspective - Structured thinking for solving data science problems.

Unit II: (Hours : 15)

Linear Algebra - Matrices and their properties (determinants, traces, rank, nullity, etc.)- Eigenvalues and eigenvectors - Matrix factorizations - Inner products; Distance measures; Projections; Notion of hyper planes- half-planes.

Unit III: (Hours: 15)

Eigenvalues and Eigenvectors - Eigenvectors and Upper Triangular matrices – Eigen spaces and Diagonal Matrices - Inner Products and Norms - Linear functionals on Inner Product spaces.

Unit IV: (Hours: 15)

Probability, Statistics and Random Processes - Probability theory and axioms; Random variables - Probability distributions and density functions (univariate and multivariate) - Expectations and moments; Covariance and correlation - Statistics and sampling distributions - Hypothesis testing of means, proportions, variances and correlations - Confidence (statistical) intervals - Correlation functions - White-noise process.

Unit V: (Hours:18)

Optimization: Unconstrained optimization; Necessary and sufficiency conditions for optima; Gradient descent methods; Constrained optimization, KKT conditions; Introduction to non-gradient techniques; Introduction to least squares optimization; Optimization view of machine learning.

BOOKS FOR STUDY:

1. G. Strang, "Introduction to Linear Algebra" , Wellesley-Cambridge Press, Fifth edition, USA, 2016.

2. Bendat, J. S. and A. G. Piersol, "Random Data: Analysis and Measurement Procedures", 4th Edition. John Wiley & Sons, Inc., NY, USA, 2010.

3. Montgomery, D. C. and G. C. Runger, "Applied Statistics and Probability for Engineers", 5th Edition, John Wiley & Sons, Inc, NY, USA, 2011.

4. David G. Luenberger, "Optimization by Vector Space Methods", John Wiley & Sons (NY), 1969.

BOOKS FOR REFERENCE:

1. Cathy O'Neil and Rachel Schutt , "Doing Data Science", O'Reilly Media, 2013.

MOOC :

- https://nptel.ac.in/courses/111/106/111106051/
- https://nptel.ac.in/courses/108/104/108104112/
- https://nptel.ac.in/courses/111/105/111105041/

WEB RESOURCES :

- https://www.cs.rutgers.edu/~cwcowan/CS674/OptimizationNotes.pdf

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Relate** the basic mathematical models and relate them in terms of data science.	**K1**
CO2	**Explain** problem solving methods for diverse data science applications	**K2**
CO3	**Utilize** the logical thinking in Context and select suitable methods.	**K3**
CO4	**Recommend** techniques related to the area of data science in variety of applications and perform in terms of mathematics..	**K4&K5**

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	M	S
CO2	S	S	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M

S - Strong, M - Medium, L - Low

Programme Title : **M.Sc. Data Science**

Course Title : **ELECTIVE- II TEXT ANALYTICS**

Course Code : **22PDSEC2**   **Hours/Week: 4**

Semester : **II**   **Credit: 4**

COURSE OBJECTIVES:
1. Provide fundamental knowledge in text mining.
2. Introduce basic concepts in text analytics
3. Familiarize various area of text analytics
4. Identify and integrate the standard business intelligence tools.

SYLLABUS

Unit - I:(Hours : 13)

History of Text Mining: Roots of text mining-Information extraction and text mining - Development of enabling technology in text mining - Sentiment analysis and opinion mining.

Unit - II :(Hours : 10)

Basics of Text Analytics : Definition - Business challenges addressed: information organization and access - Discovery of patterns- Discovery.

Unit - III: (Hours :10)

Seven Practice Areas of Text Analytics :Seven practice areas of text analytics-Finding the appropriate solution to a problem - Overall relationship - Visualizing the domains of text analytics.

Unit - IV: (Hours : 15)

Future of Text and Web Analytics: Text analytics and text mining- Future of web analytics - Future of text mining - Integration of web analytics with standard business intelligence tools- New are as that may use text analytics.

Unit - V: (Hours: 12)

Case study for text mining for modeling using python

BOOKS FOR STUDY:

1. Gary Miner John ElderIV-Robert Nisbet-Dursun Delen-Thomas Hill- Andrew Fast - "Practical Text Mining and Statistical Analysis for Non- structured Text Data Applications"-1stEdition-AcademicPress-2012.

2. BrianCliftonSybex-"AdvancedWebMetricswithGoogleAnalytics"-3rd Edition - Pearson education -2012.

MOOC :

1. https://nptel.ac.in/courses/110/107/110107129/

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Define** the important concepts of Text analytics framework	**K1**
CO2	**Relate various techniques** to pinpoint the problems and techniques for current scenario.	**K2**
CO3	**Make use of** text analytics in different areas of society and helps to identify the pattern of data.	**K3**
CO4	**Examine and interpret** python for real-world problems and provide quick solution.	**K4&K5**

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	S	M	S
CO2	S	S	M	S	S
CO3	S	M	S	M	S
CO4	S	S	M	S	M

S - Strong, M - Medium, L - Low

Program me Title : M.Sc. Data Science

Course Title   : ELECTIVE- II WEB ANALYTICS

Course Code   : 22PDSEC2A   **Hours/Week: 4**

Semester   : II   **Credit:4**

COURSE OBJECTIVES:
   1. Provide the fundamental concepts and importance of Web analytics
   2. Elucidate the role of Web analytic in collecting, analysing and reporting website traffic.
   3. Identify key tools and diagnostics associated with Web analytics.
   4. Explore effective Web analytics strategies in e-Commerce, business research, and market research

SYLLABUS

Unit- I : Introduction to Web Analytics (Hours: 12)

Introduction to Web Analytics: Web Analytics Approach- A Model of Analysis- Context matters- Data Contradiction- Working of Web Analytics: Log file analysis- Page tagging- Metrics and Dimensions- Interacting with data in Google Analytics.

Unit- II: Learning about users through Web Analytics (Hours : 10)

Goals: Introduction- Goals and Conversions- Conversion Rate- Goal reports in Google Analytics- Performance Indicators- Analyzing Web Users: Learning about users- Traffic Analysis- Analyzing user content- Click-Path analysis- Segmentation.

Unit - III: Google Analytics (Hours: 10)

Different analytical tools - Key features and capabilities of Google analytics- How Google analytics works - Implementing Google analytics - Getting up and running with Google analytics -Navigating Google analytics- Using Google analytics reports -Google metrics - Using visitor data to drive website improvement- Focusing on key performance indicators- Integrating Google analytics with third-Party applications.

Unit - IV: Overview of Qualitative analysis (Hours:10)

Lab Usability Testing- Heuristic Evaluations- Site Visits- Surveys (Questionnaires) - Testing and Experimentation: A/B Testing and Multivariate Testing-Competitive Intelligence - Analysis Search Analytics: Performing Internal Site Search Analytics, Search Engine Optimization (SEO) and Pay per Click (PPC)-Website Optimization against KPIs- Content

optimization- Funnel/Goal optimization - Text Analytics: Natural Language Processing (NLP)- Supervised Machine Learning (ML) Algorithms-API and Web data scarping using R and Python.

Unit V: Visual Analytics (Hours:18)

Case Studies : VISUAL ANALYTICS: Drill down and hierarchies-Sorting- Grouping- Additional Ways to Group- Creating Sets- Analysis with Cubes and MDX- Filtering for Top and Top N- Using the Filter Shelf- The Formatting Pane- Trend Lines- Forecasting- Formatting- Parameters - SOCIAL NETWORK ANALYSIS: Types of social network-Graph Visualization-Network Relationships-Network structures: equivalence-Network Evolution- Diffusion in networks- Descriptive Modeling-Predictive Modeling-Customer Profiling- Network targeting.

BOOKS FOR STUDY:

1. Beasley M, "Practical web analytics for user experience: How analytics can help you understand your users", Newnes, 1st edition, MorganKaufmann, 2013,.

2. Sponder M, "Social media analytics: Effective tools for building, interpreting, and using metrics", 1st edition, McGraw Hill Professional, 2013,

3. Clifton B, "Advanced Web Metrics with Google Analytics", 3rd edition, John Wiley &Sons, 2012.

BOOKS FOR REFERENCE:

1. Peterson E. T, "Web Analytics Demystified: AMarketer's Guide to Understanding How Your Web Site Affects Your Business", 2004,

2. SostreP,LeClaire J, "Web Analytics for dummies", John Wiley & Sons, 2007.

3. Burby J, Atchison S, "Actionable web analytics: using data to make smart business decisions", John Wiley & Sons, 2007.

4. 4. Dykes B, "Web analytics action hero: Using analysis to gain insight and optimize your business", AdobePress, 2011.

WEB RESOURCES :

1. https://www.selecthub.com/business-intelligence/visual-analytics-tools/
2. https://www.inc.com/guides/12/2010/11-best-web-analytics-tools.html
3. https://analytics.google.com/analytics/web/

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Relate** the numerous areas of web analytics to understand the background	K1
CO2	**Outline** sources of web data in terms of users, social media and web intelligence..	K2
CO3	**Discover** bottom-line of multiple features of analytics and its role in analysis.	K4
CO4	**Explain** diverse tools to explore user and business intelligence.	K5

Mapping of COs with POs :

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	M	S
CO2	S	S	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M

S - Strong, M - Medium, L - Low

Programme Title   :   M.Sc. Data Science

Course Title   :   CORE PRACTICAL – III VISUALIZATION LAB

Course Code   :   22PDSQC3   Hours/Week: 5

Semester   :   II   Credit: 5

Course Objectives:

1. To teach the various visualization tools.
2. To explain why visualization is an important part of data analysis
3. Use knowledge of perception and cognition to evaluate visualization design alternatives.

SPSS

1. Descriptive statistics- mean, median and standard deviation draw graphs using SPSS
2. Handling external data-files, data cleaning data and visualizing
3. Histograms and the Normal Distribution
4. Correlation Coefficient for Bivariate Data: Two Variables
5. Performing an independent-samples t-test

D3.js

1. Draw basic shapes using SVG and CSS
2. Draw Straight line & Curved lines with the SVG path
3. Draw Bar Chart, Circle Chart, Pie Chart
4. Line Chart, Bubble Chart
5. Create Animation

Tableau

1. Import external sales-data file into tabula Plot Sales against Category, Segment, and Order Date.
2. Data Blending, subset selection
3. Scatter plots and Motion charts
4. Clustering
5. Reference line, Trend line, and Forecasting

BOOKS FOR STUDY:

1. Scott Murray, "Interactive data visualization for the web: An Introduction to Designing with D3", O'Reilly, 2nd Edition, 2017
2. Joshua N. Milligan, "Learning Tableau 2019: Tools for Business Intelligence, data prep, and visual analytics", Packt Publishing Limited, 2019.

BOOKS FOR REFERENCES:

1. Ritchie S. King - Visual Storytelling with D3- An Introduction to Data Visualization with D3, Addison-Wesley-Data Analytic Series, ISBN 10: 0321933176

2. Elijah Meeks , D3.js in Action, Second Edition: Data visualization with JavaScript, Publisher: Manning Publications, 2017 , ISBN: 9781617294488
3. Ben Fry, "Visualizing Data", O "Reilly Media, Inc., 2007.

WEB RESOURCES

- https://www.crumplab.com/statisticsLab/lab-1-graphing-data.html#practice-problems
- https://www.tutorialspoint.com/d3js/d3js_drawing_charts.htm
- https://github.com/d3/d3/blob/main/API.md#color-schemes-d3-scale-chromatic
- https://www.softwaretestinghelp.com/d3-js-tutorial/#Example_Of_SVG_Transformation

COURSE OUTCOMES(CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Recall** the importance of various data visualization components	K1
CO2	**Infer** the EDA methods using data analytics tools	K2
CO3	**Apply** the data analysis methods for real time data and effectively presents the data using visualizations techniques	K3
CO4	**Analyze** appropriate design principles in the creation of presentations, critique, and revise data visualizations	K4
CO5	**Create** multiple versions of digital visualizations using various software packages	K5

Mapping of COs with POs:

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	S	S
CO2	S	S	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M
CO5	M	S	S	M	M

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title : CORE PRATICAL- IV R PROGRAMMING LAB

Course Code : 22PDSQC4   **Hours/Week: 5**

Semester : II   **Credit: 4**

COURSE OBJECTIVES:
1. Expand R by installing R packages.
2. Read Structured Data into R from various sources.
3. Understand the different data types in R.
4. Understand the R coding.
5. Analyze and visualize different Data Analytical methods in "R"

1. R AS CALCULATOR APPLICATION
   a. Using with and without R objects on console
   b. Using mathematical functions on console
   c. Write an R script, to create R objects for calculator application and save in a specified location in disk.

2. DESCRIPTIVE STATISTICS IN R
   d. Write an R script to find basic descriptive statistics using summary, str, quartile function on mtcars& cars datasets.
   e. Write an R script to find subset of dataset by using subset (), aggregate () functions on iris dataset.

3. READING AND WRITING DIFFERENT TYPES OF DATASETS
   a. Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location.
   b. Reading Excel data sheet in R.
   c. Reading XML dataset in R.

4. VISUALIZATIONS
   a. Find the data distributions using box and scatter plot.
   b. Find the outliers using plot.
   c. Plot the histogram, bar chart and pie chart on sample data

5. CORRELATION AND COVARIANCE
   a. Find the correlation matrix.
   b. Plot the correlation plot on dataset and visualize giving an overview of relationships among data on iris data.

6. REGRESSION MODEL

Import a data from web storage. Name the dataset and now do Logistic Regression to find out relation between variables that are affecting the admission of a student in a institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not. require (foreign), require(MASS).

7. **MULTIPLE REGRESSION MODEL**

   Apply multiple regressions, if data have a continuous independent variable. Apply on above dataset.

8. **REGRESSION MODEL FOR PREDICTION**

Apply regression Model techniques to predict the data on above dataset.

9. **CLASSIFICATION MODEL**

   a. Install relevant package for classification.

   b. Choose classifier for classification problem.

   c. Evaluate the performance of classifier.

10. **CLUSTERING MODEL**

   a. Clustering algorithms for unsupervised classification.

   b. Plot the cluster data using R visualizations

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Define** the syntax and semantics of R Programming	**K1**
CO2	**Demonstrate** R Studio through readings, practice exercises and writing R code.	**K2**
CO3	**Develop** of a data-mining application using sample, realistic data sets and modern tools	**K3**
CO4	**Measure** interesting patterns from different kinds of database.	**K5**
CO5	**Design** hypotheses based on the analysis to conceptualize a DM solution to a practical problem	**K6**

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	S	M	M	M
CO2	S	S	S	S	S
CO3	S	M	S	S	S
CO4	S	S	M	M	M
CO5	S	S	S	M	S

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title : CORE PRATICAL- VII BIG DATA ANALYTICS

Course Code : 22PDSC7   **Hours/Week: 5**

Semester : III   **Credit: 5**

COURSE OBJECTIVES

1. To understand the Big Data Platform and its Use cases
2. To provide HDFS Concepts and Interfacing with HDFS
3. To understand Map Reduce Jobs and apply analytics on Structured, Unstructured Data.

Unit I: Introduction to Big data (Hours: 15)

Introduction to big data: Introduction- Big Data- Characteristics of Big Data- Big data management architecture- Examining Big Data Types- Big Data Technology Components - Big data analytics- Big data analytics examples - Web Data Overview- Web Data in Action

Unit II: Hadoop (Hours: 15)

Hadoop : Introduction- History of Hadoop - Hadoop Ecosystem- Analyzing data with Hadoop - Hadoop Distributed File System- Design - HDFS concepts - Hadoop filesystem-Data flow - Hadoop I / O - Data integrity- Serialization - Setting up a Hadoop cluster - Cluster specification -cluster setup and installation- YARN

Unit III: MapReduce (Hours: 15)

MapReduce: Introduction- Understanding Map, Reduce functions - Scaling out - Anatomy of a MapReduce Job Run - Failures- Shuffle and sort - Mapreduce types and formats - features - counters - sorting - Mapreduce Applications- Configuring and setting the environment - Unit test with MR unit- local test

Unit IV: Spark (Hours: 15)

Spark:- Installing spark- Spark applications, Jobs, Stages and Tasks -Resilient Distributed databases- Anatomy of a Spark Job Run- Spark on YARN- SCALA: Introduction- Classes and objects- Basic types and operators- built-in control structures- functions and closures- inheritance

Unit V: NOSQL (Hours: 15)

NoSQL Databases: Introduction to NoSQL- MongoDB: Introduction- Data types --

Creating,Updating and deleing documents -Querying- Introduction to indexing- Capped collections. Hbase: Concepts- Hbase Vs RDBMS-Creating records-Accessing data- Updating and deleting data -Modifying data- exporting and importing data.

BOOKS FOR STUDY:

1. John Wiley &Sons ,Taming the Big Data Tidal wave, Bill Franks (2012).

2. Tom White Hadoop : The Definitive Guide, O'ReillyMedia, Third Edition (2012).

3. Martin Odersky, Lex Spoon,, Bill Venners, Programming in Scala, Second Edition, Artima Press,California(2010).

BOOKS FOR REFERENCE:

1. Boris lublinsky, Kevin t. Smith, Alexey , Yakubovich (2015). Professional Hadoop Solutions, Wiley

2. Chris Eaton, Dirk deroos et al.(2012). Understanding Big data, McGrawHill

3. Min Chen. Big Data : Related Technologies, Challenges and Future Prospects, Springer (2014).

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Identify** Big Data and its Business Implications.	K1
CO2	**Infer** the skills to handle the structured, semi structured and Unstructured data.	K2
CO3	**Apply** tools and technology in Big data.	K3
CO4	**Differentiate** various big data technologies like Hadoop MapReduce, Pig, Hive, Hbase and No-SQL.	K4
CO5	**Develop** coding over big data echo systems	K6

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	S	M	M	M
CO2	S	S	M	S	S
CO3	S	M	S	S	S
CO4	S	S	M	M	M
CO5	S	S	M	S	S

S - Strong, M - Medium, L - Low

Programme Title   :   M.Sc. Data Science

Course Title   :   CORE PAPER- VIII MACHINE LEARNING

Course Code   : 22PDSC8   **Hours/Week: 5**

Semester   : III   **Credit: 5**

COURSE OBJECTIVES

1. To Learn about Machine Intelligence and Machine Learning applications
2. To implement and apply machine learning algorithms to real-world applications.
3. To identify and apply the appropriate machine learning technique to classification, pattern recognition, optimization and decision problems.
4. To understand how to perform evaluation of learning algorithms and model selection

SYLLABUS

Uni- I: BASICS OF MACHINE LEARNING (Hours: 15)

Machine Learning- Machine Learning Foundations- Overview- Design of a Learning system- Types of machine learning -Applications Mathematical foundations of machine learning- random variables and probabilities- Probability Theory- Probability distributions- Decision Theory- Bayes Decision Theory- Information Theory.

Unit - II: SUPERVISED LEARNING   (Hours: 15)

Linear Models for Regression- Linear Models for Classification- Naïve Bayes- Discriminant Functions- Probabilistic Generative Models- Probabilistic Discriminative Models- Bayesian Logistic Regression. Decision Trees- Classification Trees- egression Trees- Pruning. Neural Networks- Feed- forward Network Functions- Back- propagation. Support vector machines- Ensemble methods- Bagging- Boosting.

Unit- III: UNSUPERVISED LEARNING (Hours: 15)

Clustering- K means- EM Algorithm- Mixtures of Gaussians. The Curse of Dimensionality- Dimensionality Reduction- Factor analysis- Principal Component Analysis- Probabilistic PCA- Independent components analysis .

Unit IV: PROBABILISTIC GRAPHICAL MODELS (Hours: 15)

Graphical Models- Undirected graphical models- Markov Random Fields- Directed Graphical Models- Bayesian Networks- Conditional independence properties- Inference- Learning- Generalization- Hidden Markov Models- Conditional random fields(CRFs) .

Unit V : ADVANCED LEARNING (Hours: 15)

Sampling- Basic sampling methods- Monte Carlo. Reinforcement Learning- K- Armed Bandit- Elements- Model -Based Learning- Value Iteration- Policy Iteration. Temporal Difference Learning- Exploration Strategies- Deterministic and Non- deterministic Rewards and Actions Computational Learning Theory- Mistake bound analysis, sample complexity analysis, VC dimension. Occam learning, accuracy and confidence boosting.

BOOK FOR STUDY:

1. Tom M. Mitchell, Machine Learning, McGraw-Hill Education (India) Private Limited, 2013.

BOOKS FOR REFERENCE :

1. EthemAlpaydin, Introduction to Machine Learning, MIT Press, Third Edition, 2014.
2. Kevin P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012
3. Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning, Springer, Second Edition, 2011.

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Recall** the fundamental concepts and theories of machine learning models	K1
CO2	**Apply** scaling up machine learning techniques, associated computing techniques and technologies	K2& K3
CO3	**Analyze** the mathematical relations across Machine Learning paradigms.	K4
CO4	**Design** various machine learning algorithms in real-world applications	K5

Mapping of COs with POs:

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	S	M	M	M
CO2	S	S	S	S	S
CO3	S	M	S	S	S
CO4	S	S	M	M	M

S - Strong, M - Medium, L - Low

Programme Title :   M.Sc. Data Science

 Course Title   :   ELECTIVE- III CLOUD COMPUTING

Course Code   : 22PDSEC3   **Hours/Week: 4**

Semester   : III   **Credit: 4**

COURSE OBJECTIVES:
1. Articulate the main concepts, key technologies, strengths and limitations of cloud computing.
2. Understand the cloud architectures and its storage.
3. Elucidate the core issues of cloud computing such as resource management and security
4. Evaluate the various tools for cloud development.

SYLLABUS

Unit I: Introduction (Hours: 12)

Introduction to Cloud Computing- Desired Features of Cloud Computing- Challenges and Risks- Benefits and Disadvantages of Cloud Computing.- Evolution of Cloud Computing- Underlying Principles of Parallel and Distributed Computing- Cloud Characteristics- Elasticity in Cloud- On-demand Provisioning.

Unit II: Cloud Enabling Technologies (**Hours : 12)**

Service Oriented Architecture- REST and Systems of Systems- Web Services- Publish-Subscribe Model- Basics of Virtualization- Types of Virtualization- Implementation Levels of Virtualization- Virtualization Structures- Tools and Mechanisms- Virtualization of CPU-Memory- I/O Devices -Virtualization Support and Disaster Recovery.

Unit III: Cloud Architecture, Services and Storage (Hours: 12)

Layered Cloud Architecture Design- NIST Cloud Computing Reference Architecture- Public, Private and Hybrid Clouds- laaS- PaaS- SaaS- Architectural Design Challenges- Cloud Storage- Storage-as-a-Service- Advantages of Cloud Storage- Cloud Storage Providers- S3.

Unit IV: Resource Management and Security in Cloud (Hours: 14)

Inter Cloud Resource Management- Resource Provisioning and Resource Provisioning Methods- Global Exchange of Cloud Resources- Security Overview- Cloud Security Challenges- Software-as-a-Service Security- Security Governance- Virtual Machine Security-IAM- Security Standards.

Unit V: Cloud Technologies and Advancements (Hours: 10)

Hadoop- MapReduce- Virtual Box — Google App Engine- Programming Environment for

Google App Engine - Open Stack- Federation in the Cloud- Four Levels of Federation-Federated Services and Applications- Future of Federation.

BOOKS FOR STUDY:

1. Kai Hwang, Geoffrey C. Fox, Jack G. Dongarra, "Distributed and Cloud Computing, From Parallel Processing to the Internet of Things", Morgan Kaufmann Publishers, 2012.
2. Rittinghouse, John W., and James F. Ransome, ―Cloud Computing: Implementation, Management and Security, CRC Press, 2017.

BOOKS FOR REFERENCES:

1. RajkumarBuyya, Christian Vecchiola, S. ThamaraiSelvi, ―Mastering Cloud Computing, Tata Mcgraw Hill, 2013.
2. Toby Velte, Anthony Velte, Robert Elsenpeter, "Cloud Computing- A Practical Approach, Tata Mcgraw Hill, 2009.
3. George Reese, "Cloud Application Architectures: Building Applications and Infrastructure in the Cloud: Transactional Systems for EC2 and Beyond (Theory in Practice), O'Reilly, 2009.

MOOC :

- https://onlinecourses.nptel.ac.in/noc21_cs14/preview

WEB RESOURCES :

- https://www.simplilearn.com/tutorials/cloud-computing-tutorial/cloud-computing-architecture
- https://www.7mileadvisors.com/Whitepaper/advancements-of-cloud-technology/

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Define** the study of Cloud Computing and its features development.	**K1**
CO2	**Summarize** cloud supporting technologies and virtualization resources.	**K2**
CO3	**List** numerous architectures of cloud, key concepts and benefits.	**K4**
CO4	**Justify** different security services and advanced Technologies in cloud.	**K5**

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	S	M	M	M
CO2	S	S	M	M	S
CO3	S	M	S	S	S
CO4	S	S	M	M	M

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title   : ELECTIVE – III DISTRIBUTED DATABASES

Course Code   : 22PDSEC3A   **Hours/Week: 4**

Semester   : III   **Credit: 4**

COURSE OBJECTIVES:
1. Understand the various aspects in Distributed Data.
2. Understand query processing and optimization in Distributed Database.
3. Management of distributed data with different levels of transparency.
4. Understand how to use database management tools in resolving dead lock situations.

SYLLABUS

Unit- I: Overview of Distributed Database Design(Hours :10)

What is Distributed Database System (DDBS), Features of DDBS, promises of DDBS, Design issue in DDBS,Distributed DBMS architecture:-Client/serverSystem,Peer-to-Peer,Multi-Database system, Levels of distribution transparency : Reference Architecture for Distributed Databases, Types of Data Fragmentation, Integrity Constraints in Distributed Databases Framework of Distributed Databases Design, Design of Database Fragmentation, Allocation of fragments, Transparencies in Distributed Database Design.

Unit- II: Distributed Query Processing And Optimization (Hours :15)

Concept, objective, and phases of distributed query processing, Translation of global queries to fragment queries, Query optimization in centralized databases, framework for query optimization in Distributed databases, join queries,general queries.

Unit- III: Transactions Management (Hours :10)

TRANSLATION OF GLOBAL QUERIES TO FRAGMENT QUERIES: Equivalence Transformations For Queries,Transforming Global Queries into Fragment Queries, Distributed Grouping and Aggregate Function Evaluation, Parametric Queries.

THE MANAGEMENT OF DISTRIBUTED TRANSACTIONS:A Framework for Transaction Management,Supporting Atomicity of Distributed Transactions, Concurrency Control for Distributed Transactions, Architectural Aspects of Distributed Transactions, Transaction Schedules in Distributed databases

Unit- IV: Concurrency Control and Reliability (Hours : 12)

CONCURRENCYCONTROL: Foundations of Distributed Concurrency Control, Distributed Deadlocks,ConcurrencyControlBasedonTimestamps,OptimisticMethodsforDistributedConcurrencyControl.Introduction to Deadlock, Distributed Deadlock prevention, avoidance, detection and recovery, Two-Phase and Three-Phase Commit Protocol. RELIABILITY: Basic Concepts, Non Blocking Commitment Protocols, Reliability and Concurrency Control, Determining a Consistent View of the Network, Detection and Resolution of Inconsistency, Check points And Cold Restart. Catalog Management in Distributed Databases, Authorization and Protection.

Unit-V: Security Aspects in DDBMS (Hours :13)

Study of a variety of attacks on the components of system (such as on routing protocols in ad hoc networks), privacy issues in Peer to Peer systems, trusted collaboration and dissemination of data among cooperative entities,Security problems,security policies, DAC methods,MAC methods,security models for DDBMS.

BOOKS FOR STUDY:

1. StefanoCeri,GiuseppePelagatti,"Distributed Databases principles &systems", 2ndedition,McGraw-Hill,NewYork,1985,ISBN0-07-010829-3.

2. N.TamerOzsu,PatrickValduriez,"PrinciplesofDistributedDatabaseSystems",2nd,IllustratedEdition,Prentice HallInternationalInc.,1999,ISBN 0136597076,9780136597070.

3. SilberschatzAndKorth,
   " DatabasesystemConcept",SilberschatzAndKorth6thEdition,TataMcgrawHill EducationPrivate Limited,ISBN- 9789332901384.

BOOKS FOR REFERENCE:

1. Thomas Connolly, Carolyn Begg,"Database Systems: A Practical Approach to Design, Implementation and Management",Pearson Publisher, 4th Edition.

2. Raghu Ramakrishnan and Johannes Gehrke ,"Database Management Systems", McGraw-Hill Educationpublisher, illustratedEdition, 2003, ISBN 0072465638, 9780072465631.

3. CarloZaniolo,StefanoCeri,ChristosFaloustsos,R.T.Snodgrass,V.S.Subrahmanian, "Advanced Database Systems", Morgan Kaufman,1997.

WEB RESOURCES :

- https://nptel.ac.in/courses/106/106/106106168/

- https://www.youtube.com/watch?v=hNOEL3SM9No

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Explain** the general framework of distributed databases	K2
CO2	**Choose** the appropriate framework for query optimization	K3
CO3	**Analyse** the different aspects of management of transactions	K4
CO4	**Elaborate** the various security methods	K6

Mapping of COs with POs:

COs	POs				
	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	M	S
CO2	S	S	S	S	S
CO3	S	M	S	M	S
CO4	S	S	S	S	S

S - Strong, M - Medium, L - Low

Programme Title : **M.Sc. Data Science**

Course Title : **ELECTIVE- IV INTERNET OF THINGS**

Course Code : **22PDSEC4** Hours/Week: 4

Semester : **III** Credit: 4

COURSE OBJECTIVES:
1. Provide fundamental knowledge in IoT.
2. Introduce standardized protocols for IoT.
3. Familiarize various design and services in IoT
4. Explain various platforms and applications of IoT.

SYLLABUS

Unit- I:(Hours : 10)

Fundamentals of iot - Evolution of Internet of Things - Enabling Technologies -IoT Architectures: oneM2M, IoT World Forum (IoTWF) and Alternative IoT models - Simplified IoT Architecture and Core IoT Functional Stack - Fog, Edge and Cloud in IoT- Functional blocks of an IoT ecosystem - Sensors, Actuators, Smart Objects and Connecting Smart Objects.

Unit-II: (Hours : 10)

Protocol Standardization for IoT- Efforts - M2M and WSN Protocols - SCADA and RFID Protocols - Issues with IoT Standardization - Unified Data Standards - Protocols - IEEE 802.15.4 - BACNet Protocol- Modbus - KNX -Zigbee- Network layer - APS layer - Security.

Unit - III: (Hours :15)

Design and Development - Design Methodology - Embedded computing logic - Microcontroller, System on Chips -IoT system building blocks - Arduino - Board details, IDE programming - Raspberry Pi - Interfaces and Raspberry Pi with Python Programming.

Unit - IV: (Hours: 15)

Data Analytics and support services - Structured Vs Unstructured Data and Data in Motion Vs Data in Rest - Role of Machine Learning - NoSQL Databases - Hadoop Ecosystem - Apache Kafka, Apache Spark - Edge Streaming Analytics and Network Analytics -Xively Cloud for IoT, Python Web Application Framework - Django - AWS for IoT.

Unit - V: (Hours: 10)

The Role of the Internet of Things for Increased Autonomy and Agility in Collaborative

Production Environments - Resource Management in the Internet of Things: Clustering, Synchronisation and Software Agents. Applications - Smart Grid - Electrical Vehicle Charging.

BOOKS FOR STUDY :

1. David Hanes, Gonzalo Salgueiro, Patrick Grossetete, Rob Barton and Jerome Henry, "IoT Fundamentals: Networking Technologies, Protocols and Use Cases for Internet of Things", Cisco Press, 2017

BOOKS FOR REFERENCE:

1. ArshdeepBahga, Vijay Madisetti, "Internet of Things - A hands-on approach, Universities Press", 2015

2. Olivier Hersent, David Boswarthick, Omar Elloumi , "The Internet of Things - Key applications and Protocols", Wiley, 2012.

3. Jan Ho¨ ller, VlasiosTsiatsis , Catherine Mulligan, Stamatis , Karnouskos, Stefan Avesand. David Boyle, "From Machine-to-Machine to the Internet of Things - Introduction to a New Age of Intelligence", Elsevier, 2014.

4. Dieter Uckelmann, Mark Harrison, Michahelles, Florian (Eds), ―Architecting the Internet of Things, Springer, 2011.

5. Michael Margolis, Arduino Cookbook, Recipes to Begin, Expand, and Enhance Your Projects, 2nd Edition, O'Reilly Media, 2011.

6. Olivier Hersent, Omar Elloumi and David Boswarthick, "The Internet of Things: Applications to the Smart Grid and Building", Wiley, 2012.

7. Honbo Zhou, "The Internet of Things in the Cloud: A Middleware Perspective", CRC Press,2012.

WEB RESOURCES:

- https://www.iotforall.com/what-is-iot-simple-explanation/

- https://www.iotforall.com/internet-of-things-examples-applications/

- https://www.i-scoop.eu/internet-of-things-guide/internet-of-things-examples-cases/

MOOC :

- https://onlinecourses.nptel.ac.in/noc21_cs17/preview

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Define** fundamentals of IoT with respect to various resources.	**K1**
CO2	**Outline** various architecture and its performance relate to standards.	**K2**
CO3	**Inspect** methodologies and services suitable for real-time applications.	**K4**
CO4	**Determine** opted programming platform for users of business and technical field.	**K5**

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	S	S
CO2	S	S	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M

S - Strong, M - Medium, L - Low

Programme Title : M.Sc. Data Science

Course Title : ELECTIVE- IV REALTIME ANALYTICS

Course Code : 22PDSEC4A Hours/Week: 4

Semester : III Credit: 4

COURSE OBJECTIVES:

1. To teach the fundamental techniques and principles in achieving data analytics with scalability and streaming capability.
2. To provide an overview of an exciting growing field of data analytics.
3. To enable students to have skills that will help them to solve complex real-world problems in decision support

SYLLABUS

Unit I: (Hours: 10)

Data Analytics Basics, Data Types, Analytics Types, Data Analytics Steps: Data Pre-Processing, Data Imputation, Data Cleaning, Data Transformation, Data Visualization, and Data Engineering. Descriptive, Predictive, and Prescriptive Analytics.

Unit II: (Hours: 10)

Data Analytics using Python, Statistical Procedures, Web Scraping in Python, Advanced analytics, NumPy, Pandas, SciPy, Matplotlib.

Unit III: (Hours: 15)

Box-Jenkins Methodology for ARIMA models: Examining correlation and stationarity of time series data, ARIMA models for time series data (An Auto-regressive model of order one and a Moving Average Model of order one).

Unit IV: (Hours: 15)

Streaming Analytics Architecture: Designing Real-Time Streaming Architectures, Service Configuration and Coordination.

Unit V: (Hours: 10)

Data-Flow Management in Streaming Analysis, Processing Streaming Data, Storing Streaming Data,

BOOKS FOR STUDY:

1. Anil Maheshwari, "Data Analytics made accessible," Amazon Digital Publication,2014.

2. Byron Ellis, "Real-Time Analytics: Techniquesto Analyze and Visualize Streaming Data",WILEY Publication, 2016.

3. Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley,2013.

BOOKS FOR REFERENCE:

1. Anil Maheshwari, "Data Analytics made accessible," Amazon Digital Publication, 2014

2. Real-Time Analytics, Techniques to analyze and visualize streaming Data by ByronEllis, 2016.

Web Resources :

1. **https://www.solvexia.com/blog/real-time-analytics**

2. **https://www.upsolver.com/blog/streaming-data-architecture-key-components**

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Match** principles, methodologies and applications of real time systems in entire concepts of data handling..	**K1**
CO2	**Illustrate** the various components of programming languages used for data analytics.	**K2**
CO3	**Categorize** various time series techniques of mathematics to handle real time data.	**K4**
CO4	**Conclude** various methods of storing and streaming of data..	**K5**

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	MM	S
CO2	S	S	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	S

S - Strong, M - Medium, L - Low

Programme Title :   M.Sc. Data Science

Course Title   :   ELECTIVE- V : BIG DATA LAB

Course Code   : 22PDSEC5   **Hours/Week: 4**

Semester   :   III   **Credit: 2**

COURSE OBJECTIVES:

1. To Implementation skill of data structure.
2. To know the installation steps of Hadoop.
3. To gain understanding about Hadoop environment.
4. To gain knowledge of data base management in Pig and Hive.
5. To gain understand the high-level of abstraction for processing over the MapReduce

1. Implement the following Data structures in Java

i) Linked Lists   ii) Stacks   iii) Queues iv) Set v) Map

2. (i) Perform setting up and Installing Hadoop in its three operating modes:

      Standalone, Pseudo distributed, Fully distributed

  (ii)Use web based tools to monitor your Hadoop setup.

3. Implement the following file management tasks in Hadoop:

     Adding files and directories

     Retrieving files

     Deleting files

(Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities.)

4. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.

5. Write a Map Reduce program that mines weather data. Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented.

6. Implement Matrix Multiplication with Hadoop Map Reduce

7. Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your

data.

8. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes

COURSE OUTCOMES (CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Identify** Big Data and its Business Implications.	**K1**
CO2	**List** the components of Hadoop and Hadoop Eco-System	**K2**
CO3	**Process** Data on Distributed File System	**K3**
CO4	**Manage** Job Execution in Distributed Environment	**K4**
CO5	**Develop** Big Data Solutions using Hadoop Eco System	**K5**

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	S	S
CO2	S	S	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M
CO5	M	S	S	S	S

S - Strong, M - Medium, L - Low

Programme Title :   M.Sc. Data Science

Course Title   :   ELECTIVE- V: MINI PROJECT

Course Code   : 22PDSEC5A   **Hours/Week: 4**

Semester   : III   **Credit: 2**

COURSE OBJECTIVES:
1. Provide knowledge of open source software's
2. Gain skill of project planning
3. Familiarize the design phase of software.
4. Implementing their idea for creating software project.
5. Ensure testing of project.

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Recall** the computing concepts.	**K1**
CO2	**Demonstrate** the various domains of open source software.	**K2**
CO3	**Construct** design phase of project.	**K3**
CO4	**Examine** various software testing methods.	**K4**
CO5	**Justify** their project during final phase of implementation.	**K5**

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	S	S
CO2	S	M	S	M	S
CO3	S	M	M	M	S
CO4	M	S	S	M	S
CO5	L	S	M	S	M

S - Strong, M - Medium, L – Lows

Programme Title : **M.Sc. Data Science**

Course Title : **CORE PRACTICAL- V: MACHINE LEARNING LAB**

Course Code : **22PDSQC5** Hours/Week: 4

Semester : **III** Credit: 2

COURSE OBJECTIVES:

1. Make use of Data sets in implementing the machine learning algorithms

2. Implement the machine learning concepts and algorithms in any suitable language of choice.

LIST OF PROGRAMME

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a.CSV file.

2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

4. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.

5. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

6. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.

7. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart

Disease Data Set. You can use Java/Python ML library classes/API.

8. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.

9. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.

10. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

COURSE OUTCOMES(CO): On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**Recall** the fundamental concepts and theories of machine learning models	**K1**
CO2	**Apply** scaling up machine learning techniques, associated computing techniques and technologies	**K2& K3**
CO3	**Discover** the mathematical relations across Machine Learning paradigms.	**K4**
CO4	**Design** various machine learning algorithms in real-world applications	**K5**

Mapping of COs with POs:

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	S	S
CO2	S	M	S	M	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M

Programme Title : M.Sc. Data Science

Course Title   :   PROJECT EXTERNAL PROJECT

Course Code   :   22PDSPC   Hours/Week: 25

Semester   :   IV   Credit: 6

COURSE OBJECTIVES:

1. Provide knowledge of the working environment.
2. Gain more about software development ideas.
3. Familiarize the concepts of Software design phases.
4. Implementing their technical skills for software project.
5. Ensure testing and running of project.

COURSE OUTCOMES (CO) :

CO Number	CO Statement	Knowledge Level
CO1	**Show** the process of software development.	**K1**
CO2	**Understanding** the software development tools and methods	**K2**
CO3	**Express**ing with the programming concepts and tools they learnt.	**K3**
CO4	**Classify** the various requirements for the project developed.	**K4**

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	S	M	S	S
CO2	S	S	S	S	M
CO3	S	M	M	S	S
CO4	M	S	M	M	M

S - Strong, M - Medium, L – Low

Programme Title : M.Com.

Course Title : EDC DATABASE MANAGEMENT SYSTEMS

Course Code : 22PDSQEDC Hours/Week: 4

Semester : III Credit: 4

COURSE OBJECTIVES:
1. Provide fundamental knowledge in Access Environment
2. Introduce basic concepts in tables and filters.
3. Familiarize various features of access environment
4. Identify the key features to create forms and reports.

SYLLABUS

Unit - I : (Hours : 12)

Data- Database Meaning- Database Management system- Purpose of Database System- View of data- Relational Databases-Database Architecture and Design- Relational Databases and MS Access.

Unit - II : (Hours : 12)

Designing a Database-Build a Database- Data Types- Properties-Primary Keys- Sort, Retrieve, Analyze Data in a Database- Determine Fields-Group Fields into Tables- Table Design View- Creating Tables-Using Design View to Create Tables -Create a Table Relationship.

Unit - III : (Hours : 12)

Working with Forms-Creating Forms- Using the Form Wizard- Modifying the Form – Data Entry and Editing using a form- Displaying the data - Finding Data Using a Form - Adding a Button -Adding a Label Advanced Forms – Add a Command Button to a Form-Create a Subform.

Unit - IV : (Hours : 12)

Microsoft Access Queries- Filter Records – Exploring Query- Simple Query- Projection Query- Selection Query-Sorting the Result- Perform Calculations in a Query- Relationships and the Relationships Tool - Create Relationships- One-to-Many - One-to-One - Many-to-Many- Using Logical Expressions -And- OR- Not - Query Operators – Like- In- Query Properties- Unique Values - Totals Query- Parameter Query- Crosstab Query- Action

Queries- Updating data using a query- SQL Select Statement- Summarizing data using a query – Creating a Macro.

Unit - V : (Hours : 12)

Working with Reports- Reports- Exploring Reports- Types- Creating a Report- Customizing Reports- Organize Report Information -Formatting and Setting Report Control Properties- Add Controls to a Report- Enhance the Appearance - Summarize Report Information- Sharing Data Across Applications-Import Data into Access-Export Data-Analyze Access Data in Excel- Export Data to a Text File -Merge Access Data with a Word Document

BOOKS FOR STUDY:
1. Microsoft Access 2013 step by step , joyce Cox and Joan Lambert, Microsoft press, 2013.
2. Microsoft Access 2013 bible, Lisa A. Bucki, John Walkenbach, FaitheWempen, Michael Alexander, Dick Kusleika, John willey& sons, 2013..

BOOKS FOR REFERENCE:
1. MOS 2016 Study Guide for Microsoft Access, John Pierce, Pearson Education, 2017.
2. Microsoft Access 2019 dummies, Laurie Ann Ulrich, Ken Cook, John Willey & sons, 2019.
3. Microsoft Access 2019 bible, Michael Alexander, Dick Kusleika, John Willey & sons, 2019.

WEB RESOURCES
1. **http://www.csun.edu/~vcact008/course/312/s12-access-lab-notes.pdf**
2. **http://www.omgroup.edu.in/downloads/files/n53671e2e8ee1a.pdf**

COURSE OUTCOMES (CO) : On Completion of course the Students will be able to

CO Number	CO Statement	Knowledge Level
CO1	**List** basics concepts, applications, data models, schemas and instances of Database.	K1
CO2	**Illustrate** the coding structure to construct queries.	K2
CO3	**Develop** knowledge to handle large volume of data in dynamic and static environment.	K3
CO4	**Create** a skill to deal with real-time applications**.**	K4
CO5	**Interpret** queries and reports based on users requirements and analyze recovery techniques.	K5

Mapping of COs with POs :

COs	POs				
	PO1	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	S	S
CO2	S	M	S	M	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M
CO5	M	S	S	M	S

S - Strong, M - Medium, L - Low

Programme Title : M.Com.

Course Title   : EDC IMAGE EDITING AND ANIMATION

Course Code   : 22PDSQEDC   **Hours/Week: 4**

Semester   : III   **Credit: 4**

COURSE OBJECTIVES:

1. Introduce the basic concepts of photo shop.

2. Perform image editing with various tools of photo shop

3. Get an insight into working with banner and poster creation.

4. Build an effective animation for any real time application

PRACTICAL SYLLABUS

1. Poster Creation.

2. Banner Creation.

3. Wall Writing.

4. Create letterhead.

5. Designing ads.

6. Changing Color.

7.   Text Masking.

8.   Business Card.

9. Simple Animation.

10. Animation for any real time application.

COURSE OUTCOMES (CO) : **On Completion of course the Students will be able to**

CO Number	CO Statement	Knowledge Level
CO1	**Understand** about photo shop Tools	**K2**
CO2	**Apply** photo editing, layer masking in text and shape.	**K3**
CO3	**Analyze** the concepts and steps for doing poster and banner creation	**K4**
CO4	**Evaluate** performance in the applicability of various features of Photoshop and flash	**K5**

Mapping of COs with POs :

PO/PSO	PO				
CO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**
CO1	S	M	M	M	S
CO2	S	M	S	S	S
CO3	S	M	M	M	S
CO4	S	S	S	S	M

S - Strong, M - Medium, L – Low