M.Sc. Data Analytics

Syllabus

AFFILIATED COLLEGES

Program Code: ***

2023 – 2024 onwards

BHARATHIAR UNIVERSITY

(A State University, Accredited with "A++" Grade by NAAC,
Ranked 21st among Indian Universities by MHRD-NIRF)

Coimbatore - 641 046, Tamil Nadu, India

Program Educational Objectives (PEOs)	
The PEOs of **M.Sc. Data Analytics** programme describe accomplishments that graduates are expected to attain within five to seven years after graduation	
PEO1	Apply terminologies and principles in problem solving adapting to applications of Mathematics, Statistics, Business and emerging computing technologies in the field of Data Analytics to conceptualize real world problems.
PEO2	Exhibit proficiency as data analytics professionals through latest technologies to business and organizations in demonstrating the ability for work efficacy
PEO3	Work and collaborate with interdisciplinary backgrounds as a part of team to address the contemporary issues with innovation
PEO4	Pursue entrepreneurship, research and higher studies associated with the program to function efficiently and effectively addressing challenging problems innovatively in the society
PEO5	Communicate effectively, recognize and incorporate societal needs and constraints in their professional endeavour
PEO6	Practice their profession as Data Analyst with high regard to ethical responsibilities.

Program Specific Outcomes (PSOs)	
After the successful completion of M.Sc Data Analytics Programme, the students are expected to demonstrate	
PSO1	Knowledge on Data Analytics Principles and Components Data Acquisition, Data Transformations, Big Data Platforms for analysis and Interpretation
PSO2	Sound Knowledge of constructing data into meaningful structures by data curation and reporting to predict and gather valuable Data Insights
PSO3	Knowledge on using Statistics, Mathematics in designing Models and Algorithms for achieving Business Objectives
PSO4	Sound Knowledge on Data Analytics, Big Data Technology Tools, Visualization, Database Management, Machine Learning and Programming for Analytics of Large scale Data to support business processes and functions
PSO5	Apply data science methods in assessing data requirements and integrating data analytic problem framework for domain specific applications
PSO6	Communicate data assumptions, analysis and insights in written and visual dashboards and articulate as data story
PSO7	Knowledge on Professional and ethical responsibility on data ownership and data privacy

Program Outcomes (POs)	
On successful completion of the M. Sc. Data Analytics program	
PO1	Apply knowledge of mathematics, statistics, science and computing appropriately to model the software applications, configure software platform and analyze real time data in heterogeneous domains.
PO2	Design a system, component or process, tools to meet desired needs within realistic constraints such as economic, environmental, social, and ethical and safety contexts
PO3	Have an ability to design, implement, evaluate, analyze, interpret complex problems and data, provide sustainable computational solutions and synthesis of information to provide valid conclusion for domains of business, healthcare, environment,.
PO4	Create, Select and apply appropriate technologies, tools, techniques for data modelling, processing of complex problems and prediction for data analysis.
PO5	Communicate effectively with the computing community, and with society, about complex computing activities by being able to comprehend and write effective reports, design documentation, demographics and make effective presentations.
PO6	Manage projects and function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO7	Understand the impact of professional analytical solutions in societal and environmental contexts and apply the knowledge for benefit of individual for sustainable development.
PO8	Recognize the need for, and prepare them to engage in independent and life-long learning in the context of technological advancements for the betterment of individuals, organizations, research community and society.
PO9	Apply ethical principles, commit to professional ethics and responsibilities and human values.
PO10	Utilize the knowledge of education in understanding of data, management principles, computing solutions to apply on one's own work, as a member and leader in a team to manage project in multidisciplinary environments and societal contexts.

BHARATHIAR UNIVERSITY : : COIMBATORE 641 046

M.Sc. Data Analytics (Affiliated Colleges)

(Effective for the candidates admitted during the academic year 2023-2024 onwards)

SCHEME OF EXAMINATIONS – CBCS PATTERN

Course Code	Title of the Course	Credits	Hours		Maximum Marks		
			Theory	Practical	CIA	ESE	Total
FIRST SEMESTER							
	Paper I: Principles of Data Science	4	4		25	75	100
	Paper II: Probability & Statistics	4	4		25	75	100
	Paper III: Data Structure, Design and Analysis of Algorithms	4	4		25	75	100
	Paper IV: Python Programming	4	4		25	75	100
	Paper V: Data Privacy & Ethics	4	4		25	75	100
	Practical I: Algorithm Lab	4	-	5	40	60	100
	Practical II: Python Programming Lab	4	-	5	40	60	100
Total		**28**	**20**	**10**	**205**	**495**	**700**
SECOND SEMESTER							
	Paper VI: Advanced Database Management Systems	4	4		25	75	100
	Paper VII: Mathematical Foundation for Machine Learning	4	4		25	75	100
	Paper VIII: Data Mining and Analytics	4	4		25	75	100
	Paper IX: Data Visualization	4	4		25	75	100
	Elective-I	4	4		25	75	100
	Practical III: Data Analytics with R lab	4		5	40	60	100
	Practical IV: Data Visualization lab	4		5	40	60	100
Total		**28**	**20**	**10**	**205**	**495**	**700**

THIRD SEMESTER							
	Paper X: Big Data Analytics Framework & Tools	4	4		25	75	100
	Paper XI: Machine Learning	4	4		25	75	100
	Paper XII: Business Analytics	4	4		25	75	100
	Elective II:	4	4		25	75	100
	Practical V: Big Data Analytics Lab	4		5	40	60	100
	Practical VI: Machine Learning Lab	4		5	40	60	100
	Practical VII: Mini Project & Viva	2		4	25	25	50
	Total	**26**	**16**	**14**	**205**	**445**	**650**
FOURTH SEMESTER							
	Project and Viva Voce	8			50	150*	200
	Total	8					200
	Grand Total	90					2250
ONLINE COURSES							
1	#SWAYAM/MOOC	2					
2	#Job Oriented Certificate Course	2					

* Project Evaluation – 100 marks & Viva Voce – 50 marks in ESE

During II or III Semester (Optional)

List of Electives

Elective – I	
1.	Social Media Mining
2.	Text Analytics
3.	Virtualization and Cloud
Elective – II	
1.	Behavioural Data Analytics
2.	Internet of Things
3.	Health Care Data Analytics
4.	Deep Learning

	PRINCIPLES OF DATA SCIENCE	L	T	P	C
Core/Elective/Supportive	Core	**4**			**4**
Pre-requisite	**Nil**	**Syllabus Version**	**2023-2024**		

Course Objectives:

The main objectives of this course are to:

1. To understand Data source evolution, data Characteristics and data processing models.
2. To understand and apply data processing architecture ,Eco System Components of Big Data Frameworks HADOOP, SPARK Map Reduce

To analyze and Build Data Science use cases for specific domain and applications.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand Data sources, generations, data formats, Data Evolution, Data from various domains	K1, K2
2	Understand Big Data Characteristics What, Why, When, Limitation of traditional approaches and models. Map Big Vs to Data Domains	K3
3	Understand Big Data	K2
4	Understand the Role of Big Data and Artificial Intelligence – Ethics – AI Applications	K2-K5
5	Analyze various domains of Big Data Characteristics, Platform, Programming Model and Design Big Data framework ecosystem, and data processing framework of domains of Marketing, Health Care and Supply Chain	K4-K5 K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	**Introduction to Data Evolution & Sources**	**12-- hours**

Data Evolution: Data Development Time Line – ICT Advancement-a Perspective – Data Growth-a Perspective – IT Components-Business Process – Landscape-Data to Data Science – Understanding data: Introduction – Type of Data: Numeric – Categorical – Graphical – High Dimensional Data — Data Classification –-Data Formats: Structured, Semi-Structured and Un-Structured – Data Sources: Time Series – Transactional Data – Biological Data – Spatial Data – Social Network.

Unit:2	**Data Science**	**12-- hours**

Data Science: Data Science - A Discipline – Data Science vs Statistics, Data Science vs Mathematics, Data Science vs Programming Language, Data Science vs Database, Data Science vs Machine Learning. Data Analytics- Relation: Data Science, Analytics and Big Data Analytics. Data Science Components – Big data technology – Data Science user- roles and skills.

Unit:3	**Big Data Towards Data Science**	**12-- hours**

Big Data: Introduction To Big Data: - Evolution – Data as Economy - What is Big Data – Sources of Big Data. – Big Data Myths - Characteristics of Big Data 6Vs – Big Data Use cases - Big data-Challenges of Conventional Systems- -– Data Processing Models – Limitation of Conventional Data Processing Approaches - Data Discovery-Traditional Approach, Big Data Technology: Big Data Exploration - Data Augmentation – Operational Analysis – 360 View of Customers – Security and Intelligence.

Unit:4	**Big Data and AI : Roles and Skills**	**12-- hours**

AI: Cognitive Computing : Learning Perceptions – Terminologies - Machine Learning – Neural Networks – Deep Learning - NLP – Speech Processing – Big Data and AI – Ethics in AI Research - Advanced Applications – AI Myths – Data Science Roles Data Scientist , Data Architect, Data Analyst – Machine Learning Engineer - Skills

Unit:5	**Data Science Use cases**	**10-- hours**

Data Science & Big Data Use cases Specifications and Discussion – Data Sources Identification – Data Types –Data Classification – Data Characteristics of Big V's – Data Science P's – Big Data Frameworks – Data Analytics Classification – Applications of AI: Domains : Customer Insights – Behavioral Analysis –- Marketing – Retails – Insurance – Risk and Security –Health care – Supply Chain Logistics

Unit:6	**Contemporary Issues**	**2 hours**
Expert lectures, online seminars – webinars		
	Total Lecture hours	**60-- hours**

Text Book(s)	
1	V. Bhuvaneswari, T. Devi, "**Big Data Analytics: A Practitioner's Approach**", Sci-Tech Publications, 2016.
2	Han Hu, Yonggang Wen, Tat-Seng, Chua, XuelongLi,"**Toward Scalable Systems for Big**",
3	Seema Acharya, Subhashini Chellappan, "Big Data and Analytics", Wiley Publication, first edition. Reprint in 2016
4	Joel Grus, **"Data Science from Scratch"**, 2nd Edition, O'Reilly Publisher, ISBN: 9781492041139, May 2019
Reference Books : EBooks	
1	SinanOzdemir, Sunil Kakade, "**Principles of Data Science**", Second Edition, [Packt]
2	David Natingga, **"Data Science for Algorithms in a Week",** Second Edition, [Packt]
3	PrabhanjanTattar, Tony Ojeda, Et al, **"Practical Data Science Cookbook"**, Second Edition, [Packt], ISBN: 9781787129627
4	Lillian Pierson, Jake Porway, **"Data Science for Dummies"**, Second Edition, John Wiley & Sons, Publishers, ISBN**:** 9781119327639, 2017
5	Field Cady, **"The Data Science Handbook"**, John Wiley & Sons, Publishers, ISBN**:** 9781119092940, 2017

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

	Course Title	**Duration**	**Provider**
1.	Python for Data Science	4 Weeks	Swayam
2.	Introduction to Data Science in Python (Free)	4 Weeks	Coursera
3.	Intro to Data Science (Free)	8 Weeks	Udacity
4.	Data Science Certification Training – R Programming	14 hours	Simlilearn
5.	Data Science with Python	15 hours	Simplilearn

Web link

1. hthttps://builtin.com/data-science
2. https://www.udacity.com/course/intro-to-data-science--ud359
3. https://www.tutorialspoint.com/python_data_science/index.htm

Mapping with Programme Outcomes

COs	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**
CO1	-	S	M	M	S	-	L	M	M	S
CO2	-	-	S	S	S	-	M	M	M	M
CO3	S	S	S	S	S	M	S	M	S	S
CO4	M	M	M	S	S	M	L	M	S	S
CO5	S	S	S	S	S	M	M	S	S	S

*S-Strong; M-Medium; L-Low

		PROBABILITY AND STATISTICS	L	T	P	C
Core/Elective/Supportive		**Core**	**4**			**4**
Pre-requisite		**Nil**	**Syllabus Version**		**2023-2024**	

Course Objectives:

The main objectives of this course are to:

1. To understand the Probability Theory
2. To understand theoretical distributions and automata theory

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	To understand the concept of matrices	K2
2	To apply the principle of probability	K3
3	To apply and analyse the statistical measures for specific domain	K3
4	To analyse regression models and apply hypothesis testing specific domain	K4
5	To design and illustrate distribution models for specific domains	K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** - Create

Unit:1	**Introduction to Set Theory**	**10-- hours**

Set Theory: Basic set operations, relations and functions, transitive closure relation, principle of mathematical induction. Matrices: Properties of determinants, inverse of a matrix, Eigen values and Cayley Hamilton theorem.

Unit:2	**Probability Theory**	**12-- hours**

Introduction to Probability Theory: Sample space and events, axioms of Probability, conditional probability, Bayes' theorem, independence of events.

Unit:3	**Descriptive Statistics**	**12 -- hours**

Frequency Distribution - Continuous or Grouped Frequency Distribution - Magnitude of Class intervals - Cumulative Frequency Distribution - Two Way Frequency Distribution - Measures of Central Tendency: Arithmetic Mean, Geometric Mean - Harmonic Mean - Median, Mode - Dispersion: Overview - Mean Deviation – Standard Deviation – Combined Standard Deviation

Unit:4	**Hypothesis Testing**	**12-- hours**

Correlation –Rank Correlation –Regression -Test of Hypothesis –Large Sample Test-Small Sample Test-t test –Chi Square Test –F-test – ANOVA-one way and two way

Unit:5	**Theoretical Distribution**	**12-- hours**

Theoretical Distribution: Binominal Distribution - Obtaining Coefficient - Poisson Distribution - Normal Distribution - Poisson - Cumulative Poisson Process and its generalization - applications in different business domain - ARMA and ARIMA - Monte Carlo Simulations

Unit:6	**Contemporary Issues**	**2 --hours**

Expert lectures, online seminars – webinars

	Total Lecture hours	**60-- hours**

Text Book(s)

1	William A. R. Weiss **"An Introduction to Set Theory"** Publisher: University of Toronto 2008
2	RafVandebril, Marc Van Barel, Nicola Mastronardi, **"Matrix Computations and Semiseparab Matrices: Eigenvalue and Singular Value Methods"**, JHU Press, 2009.
3	By Vijay K. Rohatgi, A.K. Md. EhsanesSaleh. **"An Introduction To Probability And Statistics"**, ISBN: 978-1-118-79964-2, 3rd Ed , 2015.
4	S.P Guptha **"Statistical Methods"**, Sultan Chand and Sons
5	R.S.N. Pillai, Bagavathi, "Statistics Theory and Practice, S.Chand& Company, 2013

Reference Books

1	Charles E. Roberts, Jr, **"Introduction to Mathematical Proofs A Transition to Advanced Mathematics"** Denny Gulick, 4th Edition, Published by Pearson, ISBN:9780134746753, 2018.
2	John R. Hauser, **"Numerical Methods for Nonlinear Engineering Models"**, Springer Netherlands, ISBN: 9401777071, 9789401777070, 1013 pages, 2017.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

	Course Title	Duration	Provider
1	Advanced Probability Theory	12 Weeks	Swayam
2	Discrete Mathematics	12 Weeks	Swayam
3	Numerical Methods And Simulation Techniques For Scientists and Engineers	8 weeks	Swayam
4	Theory of Automation	8 Weeks	Swayam

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	M	M	M	M	M	S	S	L	S
CO2	S	M	M	M	M	M	S	S	L	S
CO3	S	S	S	S	M	M	M	M	L	S
CO4	S	S	S	S	M	M	M	M	L	S
CO5	S	S	S	S	S	M	S	S	L	S

***S-Strong; M-Medium; L-Low**

		DATA STRUCTURES, DESIGN AND ANALYSIS OF ALGORITHMS	L	T	P	C
Core/Elective/Supportive		**Core**	**4**			**4**
Pre-requisite		**Nil**	**Syllabus Version**		**2023-2024**	

Course Objectives:

The main objectives of this course are to:

1. Enable the students to learn the Elementary Data Structures and algorithms.
2. To understand and analysis concepts of Algorithmic analysis and algorithm approaches.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Develop and understand on data structures, the information arranged in memory of computer, information manipulation with the use of algorithms in a data structure.	K1, K2
2	Get knowledge about algorithms and determines their time complexity.	K2
3	Demonstrate specific search and sort algorithms using divide and conquer technique.	K3,K4
4	Gain good understanding of Greedy method and its algorithm	K2,K3
5	Design and analysis of algorithms for problem solving by using the suitable algorithmic technique	K4,K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	**Introduction to Data Structures**	**12-- hours**

Elementary Data Structure: Stack – Queue– Tree – Binary Tree – Binary Search Tree – Heap - Graphs – Graph Traversal & Search Technique.

Unit:2	**Introduction to Algorithms**	**12-- hours**

Algorithm Definition and Specification – Space complexity-Time Complexity- Asymptotic Notations: Big Oh notation – Omega notation – Theta notation – Average case analysis.

Unit:3	**Design of Algorithms**	**10-- hours**

Divide and Conquer: General Method, Binary Search, Merge sort, Quick sort.

Unit:4	**Greedy Method**	**12-- hours**

Greedy Method: Knapsack problem, Minimum Cost spanning trees, Single source shortest paths.

Unit:5	**Dynamic Programming**	**12-- hours**

Dynamic Programming: Multistage graphs, All pair's shortest paths, Travelling salesperson problem. Back Tracking: 8-queens problem, Sum of subsets, Graph coloring, Hamiltonian cycles. Branch and Bound: General method, Travelling salesperson problem.

Unit:6	**Contemporary Issues**	**2 hours**

Expert lectures, online seminars – webinars

	Total Lecture hours	**60-- hours**

Text Book(s)	
1	Ellis Horowitz, SartajSahni, SanguthevarRajasekaran, "Fundamentals of Computer Algorithms", Galgotia Publications, 2011.
2	Luciano Ramalho, "Fluent Python: Clear, Concise, and Effective Programming",*O'Reilly,* ISBN: 9781491946008, *2014.*
3	Kleinberg and Tardos: "Algorithm Design",Pearson, ISBN: 0132131080 2018.
Reference Books : EBooks	
1	Dr. Basant Agarwal, "Hands-On Data Structures and Algorithms with Python: Store, manipulate, and access data effectively and boost the performance of your applications",Packt Publishing Limited, 3rd edition, 2022
2	Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, "Introduction to Algorithms", Massachusetts Institute of Technology, MIT Press, III Edition, 2009.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]			
	Course Title	**Duration**	**Provider**
1.	Python for Absolute Beginners	4 hours 24 m	Udemy
2.	Data Structures	5 hours	Coursera
3.	Data Structures Fundamentals (Free)	6 Weeks	edX
4.	Design and Analysis of Algorithm (Free)	11 Weeks	NPTEL
5.	Design and Analysis of Algorithms (Free)	8 Weeks	SWAYAM
Web link			
1. https://www.tutorialspoint.com/design_and_analysis_of_algorithms/ 2. https://www.javatpoint.com/daa-tutorial 3. http://openclassroom.stanford.edu/MainFolder/CoursePage.php?course=IntroToAlgorithms			

Mapping with Programme Outcomes										
COs	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**
CO1	S	M	M	S	-	M	-	L	-	L
CO2	M	S	S	S	L	M	-	M	-	L
CO3	S	S	S	S	-	M	L	L	-	M
CO4	S	S	S	S	-	M	L	L	-	M
CO5	S	S	S	S	-	M	S	S	M	S

*S-Strong; M-Medium; L-Low

		PYTHON PROGRAMMING	L	T	P	C
Core/Elective/Supportive		Core	4			4
Pre-requisite		Nil	Syllabus Version		2023-2024	

Course Objectives:

1. To understand the basics of Python Data structures and Programming constructs
2. To understand and Apply Python Libraries for Data Science and Machine Learning
3. To understand and apply Exploratory Data Analytics using Data Visualization

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand the basic programming structure-List, Dictionary, Tuple, String	K1,K2
2	Understand the Control structures and object oriented concepts	K1,K2
3	Design and Analyze dataset applying statistical models, visualization and models using various tools	K3,K4
4	Understand the visualization methods , packages, statistical packages and other packages for building data models	K3,K4, K6
5	Design data analytic model using the packages in python and provide inferences for multi-disciplinary domains	K3,K4

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	Introduction	10-- hours

Introduction to Python: Python Introduction, History of Python, Python features , Python interpreter, Overview of programming in Python, Basic data types, Program input and Program output, Variables and assignment. Global and local variables. Python - Basic Operators: Arithmetic Operators, Comparison Operators, Logical (or Relational) Operators, Assignment Operators, Conditional (or ternary) Operators. Modules: Importing module, Math module Random module, Packages, Composition.

Unit:2	Advanced Data Types	12-- hours

Python Strings and string manipulation [Assigning values in strings, String manipulations, String special operators, String formatting operators, Triple Quotes, Raw String, Unicode String, Build-in-String methods], Python List : Introduction, Accessing values in list, List manipulations, List Operations, Indexing, slicing & matrices. Python Dictionary - Introduction, Accessing values, Properties, Functions in Dictionary. Python Tuples: Introduction, Operation, Accessing, Function and methods in tuples and Data Type Conversion Python sets

Unit:3	Control Structures	12-- hours

Conditional Statement: Branching (if, else-if, nested), Looping: while statement, for statements, Control Statements: break, continue and pass Statements. Python Exception Handling: Try, Catch, Finally. Functions : Defining a function , Calling a function ,Types of functions , Function Arguments Anonymous functions, Regular expressions : Match function, Search function , Modifiers. Python OOPs: Class, Object, Inheritance and Constructor.

Unit:4	Python Libraries for Data Science	12-- hours

Reading and Writing CSV Files in Python using CSV Module, NumPy [Arrays and matrices]: N-dimensional data structure, Creating array, Indexing array, Reshaping, Vectorized operations, Pandas [Data Manipulation]: Create Data Frame, Combining Data Frames, Summarizing, Columns selection, Rows selection (basic) , Rows selection (filtering) , Sorting, Descriptive statistics, Rename values, Dealing with outliers.SciPy Introduction, Basic functions, Special functions(scipy.special), Integration(scipy.integrate), Optimization (scipy.optimize).TensorFlow: Computation with TensorFlow, Regression with Tensorflow

Unit:5	Python Libraries for NLP and Visualization	12-- hours

NLTK,: tokenizing, part-of-speech tagging, stemming,Sentence Segmentation, Methods for cleaning and normalizing text.Textblobn-grams, Parsing, Spelling correction. Visualization libraries : matplotlib, Seabon: Simple Line Plots, Simple Scatter Plots, Density and Contour Plots, Histograms, Customizing Colorbars, Subplots, Text and Annotation, Visualization with Seaborn

Unit:6	Contemporary Issues	2—hours

Expert lectures, online seminars – webinars

	Total Lecture hours	60-- hours

Text Book(s)

1	Jake VanderPlas, “Python Data Science Handbook” O'Reilly, 1st Edition, 2017.
2	Andreas C. Muller & Sarah Guido “Introduction to Machine Learning with Python”, O'Reilly, 1 Edition, 2016.
3	Dr. Charles Russell Severance, Sue Blumenberg, Elliott Hauser, Aimee Andrion“Python for Everybody: Exploring Data in Python 3”,CreateSpace, 2016.

Reference Books

1	Wesley J. Chun , “Core Python Programming”, 2nd Edition, Pearson Education,2016.
2	Mark Summerfield ,“Programming in Python 3”, Pearson Education,2018.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1	PYTHON - A to Z Full Course for Beginners, https://www.udemy.com/
2	Python for Data Science, https://swayam.gov.in/
3	Python for Data Science and Machine Learning Bootcamp, https://www.udemy.com/
4	Introduction to Python Programming, https://www.udacity.com/

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	L	M	S	M	-	-	-	-	-	L
CO2	S	-	-	S	-	-	M	L	-	L
CO3	M	S	S	M	L	L	M	L	-	-
CO4	S	M	S	L	-	-	-	S	M	M
CO5	S	S	S	-	-	M	-	L	S	S

*S-Strong; M-Medium; L-Low

Course code		DATA PRIVACY AND ETHICS	L	T	P	C
Core/Elective/Supportive		Core	4			4
Pre-requisite		Nil	Syllabus Version		2023-2024	

Course Objectives:

The main objectives of this course are to:

1. To understand Data source evolution, data exploration, data format and structure.
2. To understand the importance of Data privacy, ethics and access
3. To analyse data for bias and credibility

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand Data foundation, generations, data formats, Data Evolution, Data from various domains	K1, K2
2	Understand Data privacy, ethics, importance of data ethics, Data security	K3
3	Understand Data Integrity , credibility , Features and issues of data ethics – ethical use of data	K2
4	Understand the role of metadata management – Database security – access of different data sources – Data Integration	K2-K5
5	Analyze ethical toolkits, Platform, Design and Data model, and data analytics network, principle of AI ethics	K4-K5 K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6**– Create

Unit:1	Introduction to Data Exploration	10 hours

Data foundation with data types and structures – Data Collection - Data formats – Types of data – Structured vs Unstructured – The structure of data – Data pre-processing - Check data on bias – credibility – privacy – ethics – pipeline of data access – extract – filter – sort data – Data organizing – protecting data - Data modelling techniques - Data Transformation - Data for exploration — Data analysis - Data design - Data governance

Unit:2	Data Risks and Privacy	12 hours

Data access and analysis – Risk mitigation – Risks, Harms and Benefit assessment – Sensitive data – Sensitive contexts – Data security – Data Retention – Data Minimization – Data Quality – Open data transparency – Data accountability – Introduction to Data privacy– History of privacy – Degrees of privacy – Modern privacy risks – Anonymity – Data validity – Choice of Attributes and Measures – Errors in Data Processing – Errors in Model Design – Algorithmic Fairness

Unit:3	Data Ethics with Unbiased and Objective Data	12 hours

Data Ethics – the importance of data ethics — Data anonymization – The ethical use of data - Data Science needs ethics – Data ownership – Data Integrity – Biased and Unbiased data – Fairness – Accountability – Transparency - Data credibility – Data ethics and privacy – Data anonymization – The ethical use of data – Ownership - key issues in Data ethics - Open data usage – Features and characteristics - Legal compatibility of fairness

Unit:4	**Database Security and Analytics**	**12 hours**
Relational databases - Database features – Metadata – Importance – Descriptive and structural metadata – Schemas - Metadata management – Internal and External sources – combine data – Data Integration - Access of different data sources – sorting – filtering – Large datasets – Big Query – organize and secure data		
Unit:5	**Ethics and Data Protection**	**12 hours**
Personal Data definition – Transparency – Anonymization – Physical and IT security – Procedures – Passing data to third party – Receiving data – organizing and protecting data – balancing security and analytics – Data protection – Privacy laws – Design privacy – Principles – Compliance with laws and standards – Data sharing		
Unit:6	**Contemporary Issues**	**2 hours**
Expert lectures, online seminars – webinars		
	Total Lecture hours	**60 hours**

Text Book(s)

1	G.E. Kennedy, Data Privacy Law A Practical Guide to the GDPR, 2019
2	Mike Loukides, Hilary Mason, DJ Patil, “Ethics and Data Science”, O’REILLY Media, Inc., 2018.
3	Journal Joseph, Data & Analytics 4.0, The future of work, Privacy and Trust in the Age of Artificial Intelligence, 2019.

Reference Books

1	Data Privacy, Ethics and Protection Guidance note on Big Data For Achievement of the 2030 Agenda, United Nations Development Group

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

	Course Title	Duration	Provider
1	Data Science Ethics	4 Weeks	Coursera
2	Prepare Data for Exploration	4 Weeks	Coursera
3	Introduction to data ethics – Bias, Credibility, privacy	4 Weeks	Coursera
4	Solve Business problems with AI and Machine Learning	4 Weeks	CNX

Web Links

1	Tene, Omer and Polonetsky, Jules. "Privacy in the Age of Big Data: A Time for Big Decisions." February 2, 2012. 64 Stan. L. Rev. Online 63. http://www.stanfordlawreview.org/online/privacy-paradox/big-data
2	Noam, Eli. "Privacy and Self-Regulation: Markets for Electronic Privacy." 1997. http://www.citi.columbia.edu/elinoam/articles/priv_self.htm
3	Congressional Research Service, Data Protection Law: An overview, March 25, 2019

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	-	-	M	M	S	S
CO2	S	S	S	M	M	M	S	S	S	M
CO3	S	S	S	S	S	S	M	M	M	M
CO4	S	S	S	S	M	M	M	-	-	M
CO5	M	M	S	S	S	S	-	-	S	S

*S-Strong; M-Medium; L-Low

Course code		PRACTICAL I : ALGORITHM LAB	L	T	P	C
Core/Elective/Supportive		Core			5	4
Pre-requisite		Basic Programming language	Syllabus Version		2023 Onwards	

Course Objectives:

The main objectives of this course are to:

1. This course covers the basic data structures like Stack, Queue, Tree,List.
2. This course enables the students to learn the applications of the data structures using various techniques
3. It also enable the students to understand the Algorithm Design Techniques.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Implement data structures like Stack, Queue, Tree, List using python	K4,K5
2	Understand Different Data Structures of python.	K2,K3
3	Implement Divide and Conquer technique for Sorting, Searching	K3,K4
4	Design algorithms for problem solving by using suitable algorithm techniques	K3,K4,K5,K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

LIST OF PROGRAMS — **75 hours**

1) Write a program to perform various operations on stack.
2) Write a program to perform various operation in queue.
3) Write a program to solve the tower of Hanoi problem.
4) Write a program to sort an array of an elements using quick sort.
5) Write a program to search for an element in a tree using divide & conquer strategy.
6) Write a program to solve number of elements in ascending order using Merge sort
7) Write a program to solve the knapsack problem using greedy method
8) Write a program to solve travelling salesmen problem
9) Write a program to place the 8 queens on an 8X8 matrix so that no two queens Attack.

Expert lectures, online seminars – webinars

Total Lecture hours — **75 hours**

Text Books

1	Goodrich, “Data Structures & Algorithms in Java”, Wiley 3rd edition.
2	Skiena,”The Algorithm Design Manual”, Second Edition, Springer , 2008

Reference Books

1	AnanyLevith,”Introduction to the Design and Analysis of algorithm”, Pearson EducationAsia, 2003.
2	Robert Sedgewick,Phillipe Flajolet,”An Introduction to the Analysis of Algorithms”, Addison-Wesley Publishing Company,1996.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1	https://onlinecourses.nptel.ac.in/noc19_cs48/preview
2	https://nptel.ac.in/noc/courses/noc16/SEM2/noc16-cs19/

3	https://www.tutorialspoint.com/object_oriented_analysis_design/ooad_object_oriented_analysis.htm
Course Designed By:	

Mapping with Programming Outcomes										
COs	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**
CO1	S	S	M	S	-	-	L	-	-	L
CO2	M	M	M	L	-	-	-	-	-	-
CO3	S	S	M	S	-	-	L	M	-	S
CO4	S	S	M	S	-	-	L	M	-	S

*S-Strong; M-Medium; L-Low

Course code		PRACTICAL II : PYTHON PROGRAMMING LAB	L	T	P	C
Core/Elective/Supportive		Core			5	4
Pre-requisite		Basics of any OOPs Language	Syllabus Version		2023-2024	

Course Objectives:

The main objectives of this course are to:

1. This course presents an overview of elementary data items, lists, dictionaries, sets and tuples
2. To understand and write simple Python programs
3. To Understand the OOPS concepts of Python
4. To develop web applications using Python

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand and able to write programs in Python using OOPS concepts	K2,K3
2	Implement lists, dictionaries, sets and tuples as programs	K3,K4
3	Able to write programs using Python libraries	K3,K4
4	Implement of Python libraries for Visualization	K3,K4,K5

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

LIST OF PROGRAMS	**75 hours**

Implement the following in Python:

1. Programs using elementary data items, lists, dictionaries and tuples
2. Programs using OOPS concepts
3. Programs using Python libraries (Numpy,Panda)
4. Programs using Tensorflow
5. Programs using libraries for visualization(matplotlib,lineplot, scatterplot)

Total Lecture hours	**75 hours**

Text Books

1	Bill Lubanovic, “Introducing Python”, O’Reilly, First Edition-Second Release, 2014.
2	Mark Lutz, “Learning Python”, O’Reilly, Fifth Edition, 2013.

Reference Books

1	David M. Beazley,“Python Essential Reference”, Developer’s Library, FourthEdition,2009.
2	SheetalTaneja,Naveen Kumar,”Python Programming-AModular Approach”,PearsonPublications.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1	https://www.programiz.com/python-programming/
2	https://www.tutorialspoint.com/python/index.htm
3	https://onlinecourses.swayam2.ac.in/aic20_sp33/preview

Mapping with Programming Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	-	L	L	L	-	L
CO2	S	S	S	M	-	L	L	-	-	L
CO3	S	S	S	M	-	M	L	L	-	L
CO4	S	S	S	S	-	M	L	L	-	L

*S-Strong; M-Medium; L-Low

		ADVANCED DATABASE MANAGEMENT SYSTEMS	L	T	P	C
Core/Elective/Supportive		Core	**4**			**4**
Pre-requisite		**Nil**	**Syllabus Version**		**2023-2024**	

Course Objectives:

The main objectives of this course are to:

1. To understand the concepts of DBMS, Data Model and Normal forms.
2. To understand the concepts of concurrency control and Recovery.
3. To understand basics of SQL and NoSQL databases.
4. To understand and apply MongoDB (NoSQL) for Data Analysis using CURD and User Management, and to impart knowledge on Graph Databases

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand the structure and model of the relational database management systems.	K2
2	Understand the concepts of transaction management and SQL, NoSQL database models	K3
3	Understand and create database models using MongoDB and Graph Database	K4
4	Apply MongoDB operators to retrieve data from document data stores	K3
5	Understand and apply concepts of data management indexing techniques for specific applications	K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	**Database Overview**	**10 hours**

Introduction - Database concepts, Basic components of DBMS, sources of data - data models – hierarchical – network – XML and Stores - Relational Database Design: Anomalies in a Database– Functional Dependency – Lossless Join and Dependency – Preserving Decomposition – Third Normal Form– Boyce Codd Normal Form – Multi-valued Dependency – Fourth Normal Form – Join Dependency – Project Join Normal Form –Domain Key Normal Form - SQL: Data Definition – Data Manipulation – Integrity Constraints–Views–PL/SQL.

Unit:2	**NoSQL**	**12 hours**

Indexing and Hashing – Query Processing – Transaction Processing – Concurrency Control and Recovery - Advanced Database Concepts and Emerging Applications: Distributed Databases – Object Oriented Databases - Object Relational Databases- SQL–NoSQL Tradeoffs–CAP Theorem– Eventual Consistency - NoSQL–database types – Document Oriented – Columnar – Graph – Key-Value Pair - NoSQL database, design for performance / quality parameters, documents and information retrieval.

Unit:3	**MongoDB Introduction**	**12 hours**

MongoDB- Introduction – MongoDB – Need – MongoDB Vs. RDBMS – MongoDB - MongoDB Server Configuration – Import and Export –- Data Extraction Fundamentals - Intro to Tabular Formats - Parsing CSV - Parsing XLS with XLRD-Parsing XML - Intro to JSON - MongoDB-CURD Operations – MongoDB Operators - Query Document - Pipeline - Aggregation Operators -

Unit:4	Advanced MongoDB	12 hours
User Management – MongoDB Data Replication in Servers – Data Sharding – MongoDB Indexes – Create – Find – Drop – Backup – MongoDB – Relationships – Analyzing Queries – MongoDB Objectid – Advanced MongoDB: MapReduce – MongoDB - Text Processing - Regular Expression.		
Unit: 5	**Graph Database**	**12 hours**
Introduction to graphs – Graph Database – Indexes – Graph – Nodes – Properties –Relationships – Traversal – Path - Graph Compute Engines – The power of graph databases –Performance – Flexibility – Agility - Graph Data Modeling – Types of Graphs – Non directed graphs – Directed Graphs – Weighted Graphs - Labeled Property - Graph Model – Querying Graphs – Cypher – Comparison of Relational and Graph Modeling – Building graph database application –Graph storage databases – Graph store –: Neo4j – Hyperbase – DB – InfoGrid -Graphs in the real world.		
Unit:6	**Contemporary Issues**	**2 hours**
Expert lectures, online seminars – webinars		
	Total Lecture hours	**60-- hours**

Text Book(s)

1	Abraham Silberchatz, Henry K.Forth, Sudharshan, **"Database system Concepts"**, 7th edition, McGraw Hill, 2020.
2	Prabu C.S.R, **"Object-Oriented Database Systems: Approaches and Architectures"** 3rd Edition, PHI, 2011.
3	Kristina Chodorow , **"MongoDB: The Definitive Guide"**, 3rd Edition , O'Reilly Media, ISBN: 9781491954461, 2019.
4	Guy Harrison, **"Next Generation Databases: NoSQL, NewSQL, and Big Data"**,Apress, 2016.

Reference Books :EBooks

1	ShamkantB.Navathe, RamezElamsri**"Fundamentals of Database Systems"**, 7th Edition, Pearson Education Limited, 2017.
2	David Hows , Peter Membrey , EelcoPlugge , Timm Hawkins , **"The Definitive Guide to MongoDB"**, 3rd Edition, Apress, 2015.
3	GauravVaish ,**"Getting Started with NoSQL"**Packt Publishing, 2013.
4	Ian Robinson, Jim Webber & Emil Eifrem, "Graph Databases New Opportunities for Connected Data",2ndEdition, O'Reilly publication.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

	Course Title	Duration	Provider
1.	Database Management System	12 Weeks	Swayam
2.	Database Management System	8 Weeks	NPTEL
3.	NoSQL Systems	4 Weeks	Coursera
4.	Introduction to MongoDB	3 Weeks	Coursera

Web link

1. https://www.w3schools.in/dbms/
2. https://www.guru99.com/nosql-tutorial.html
3. https://www.tutorialspoint.com/mongodb/index.htm

Course Designed by: **Dr.S.Gavaskar**

Mapping with Programme Outcomes										
COs	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**
CO1	S	M	M	S	M	M	M	M	L	S
CO3	S	M	M	S	M	M	L	M	L	M
CO3	S	M	M	S	M	M	L	M	L	M
CO4	S	M	S	S	M	S	L	M	L	S
CO5	S	M	S	M	M	M	M	M	L	S

*S-Strong; M-Medium; L-Low

		MATHEMATICAL FOUNDATIONS FOR MACHINE LEARNING	L	T	P	C
Core/Elective/Supportive		Core	**4**			**4**
Pre-requisite		**Nil**	**Syllabus Version**		**2023-2024**	

Course Objectives:

The main objectives of this course are to:

1. To understand linear programming methods.
2. To understand Dynamic programming approach.
3. To understand concepts basics concepts of Linear Algebra
4. To understand concepts of vector spaces and matrices
5. To understand the applications of Linear Algebra in Machine Learning

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Solve linear programming techniques to optimization problems arising in all Computer fields	K3
2	Use Dynamic programming approach to real time problems.	K3
3	Understand the basics of Linear Programming constructs	K2
4	Apply vector spaces and their applications in Machine Learning	K3
5	Understand the concepts of matrix, Gaussian Elimination and differential equations and Apply the concepts of Linear Algebra in Machine Learning Algorithms	K2, K5

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	**Linear Programming Problem**	**12 hours**

Introduction to Operations Research: Basics definition - scope – objectives - phases - models - limitations of Operations Research - Linear Programming Problem - Formulation of LPP - Graphical solution of LPP - Simplex Method - Artificial variables - Big-M method - Two-phase method - Degeneracy - Unbounded solutions- Duality in Linear Programming Problems – Dual Simplex

Unit:2	**Dynamic Programming**	**12 hours**

Dynamic Programming - Introduction - Characteristics of dynamic programming – Dynamic programming approach for Priority Management employment smoothening – capital budgeting – Stage Coach/Shortest Path – cargo loading and Reliability problems.

Unit:3	**Geometry Linear Equations and Vector Spaces**	**12 hours**

The Geometry of Linear Equations - An Example of Gaussian Elimination- Matrix Notation and Matrix Multiplication - Triangular Factors and Row Exchanges- Inverses and Transposes.

Vector Spaces and Subspaces – Solving Ax=0 and Ax=b - Linear Independence, Basis and Dimension- The Four Fundamental Subspaces- Graphs and Networks- Linear Transformations.

Unit:4	**Determinants, Eigen values and Eigenvectors**	**12 hours**

Determinants: Introduction- Properties of the Determinant- Formulas for the Determinant – Applications of Determinants. Eigen values and Eigenvectors: Introduction- Diagonalization of a Matrix .- Difference Equations and Powers A k- Differential Equations and e At - Complex Matrices- Similarity Transformations – A - Applications of Machine Learning – Use cases.

Unit:5	Positive Definite Matrices	10 hours
Minima, Maxima, and Saddle Points - Tests for Positive Definiteness - Singular Value Decomposition – Machine Learning Applications – Use cases.		

Unit:6	Contemporary Issues	2 hours
Use Linear and Dynamic programming approach to real time problems. Apply the concepts of Linear Algebra in Machine Learning Algorithms Expert lectures, online seminars – webinars		
	Total Lecture hours	**60 hours**

Text Book(s)

1	J K Sharma, **"Operations Research Theory &Applications"** 6th Edition, Laxmi Publications, 2017.
2	Gilbert Strang, Linear Algebra and Its Application, 5thEdition, Wellesley Cambridge Press, ISBN: 9780980232776, 2017.

Reference Books : EBooks

1	P. K. Gupta and D. S. Hira, **"Operations Research"**, S. Chand & co., 2017
2	David C. Lay, Steven R. Lay, Judi J. McDonald, **"Linear Algebra and Its Applications"** 5th Edition, Pearson Education, 2016.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

	Course Title	Duration	Provider
1.	Operations Research	15 Weeks	Swayam
2.	Linear Algebra	12 Weeks	Swayam
Web link			
1.	https://stemez.com/subjects/science/1HOperationsReseach/1HOperationsReseach.php		
2.	https://www.khanacademy.org/math/linear-algebra		

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	M	S	M	M	M	S	S	M	S
CO2	S	S	S	S	M	M	S	S	M	S
CO3	M	M	M	S	S	M	M	S	L	S
CO4	S	M	S	S	S	M	S	S	L	S
CO5	M	M	M	S	S	M	M	S	L	S

*S-Strong; M-Medium; L-Low

Course Code		DATA MINING AND ANALYTICS	L	T	P	C

Core/Elective/Supportive	Core	Semester	4
Pre-requisite	Nil	Syllabus Version	2023-2024

Course Objectives:

The main objectives of this course are to:

1. To understand the concepts of Data Warehouse architecture and apply for various domains.
2. To understand Data Mining techniques Cluster, Classification and Association Rule Mining.
3. To understand the concepts of Web mining, Text mining and Spatial mining.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand Data warehousing architecture and process.	K2
2	Apply the mining techniques like association, classification and clustering on datasets	K3
3	Understand the visualization package R	K1.K2
4	Analyse the data set to understand the issues in the real world problem	K4,K5 K6
5	Apply the statistical measures in R	K3,K5

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6**– Create

Unit:1	Data Warehousing	8 hours

Data Warehouse -Introduction - Multidimensional data model - OLAP operations - Warehouse schema - Data warehousing architecture - Warehouse Schema - Warehouse server - Meta data - OLAP Engine - Data warehouse backend process - Data Warehouse Technology - Warehousing Software - Cloud data warehousing. Data Warehousing Case Study: Government, Tourism and Industry.

Unit:2	Data Mining	12 hours

Data Mining: Introduction – Data as a Subject - Definitions- KDD vs. Data mining- Data Mining techniques-Current Trends in Data Mining. Association Rules: Concepts- Methods to discover Association rules- A priori algorithm.

Clustering: Data Attribute Types – Data Similarity and Dissimilarity - Clustering paradigms– Partition algorithm-K-Means algorithm,

Classification: Introduction – Decision Trees: Tree Construction Principle – Attribute Selection measure – Tree Pruning - Decision Tree construction Algorithm – CART – ID3.

Unit:3	Exploratory Data Analytics: Visualization Package R	14 hours

Introduction - Overview and History of R - Data Types - R Objects and Attributes - Vectors – Removing Missing Values-Combining Variables - Vectorized Operations – Apply() family - **Cleaning Data**: – Exploring Raw Data - Visualising Distributions - Typical Values - Unusual Values-Missing Values: Zeros And Nas - Filling Missing Values – Data Manipulation using dplyr() package-**Visualization Packages** – Understanding Plots - Aesthetics - Lattice – Ggplot2 – Plotly - Univariate Visualization: Histogram – Box Plot- Bar Chart - Multivariate Visualizations: Scatter Plot- Heat Map-**Reports & Dashboards:** Rmarkdown Package - Dashboards: Flex Dashboard: Layout: Row-based layouts - Attributes on sections - Multiple pages - Story boards – Components: Value boxes – Gauges-Text annotations - Navigation bar – Shiny Web App: Introduction Shiny - Layout - Control widgets-Reactive output - R scripts and Data - Reactive expressions – App Deployment.

Unit:4	Data Insights	12 hours

Data Insights: Data types – Categorical – Binary – ordinal – Nominal – continuous – Discrete Data Dimensions –Numerical Measures – Central Tendency – Mean – Median – Mode - Understanding data using central tendency – plotting histogram – density plots and inference of plot - Variability Measure – Variance - Range - IQC - and Standard Deviation – Sum of squares – Squared Deviations – Absolute Deviations - Identify outlier using Inter Quartile Range – Visualization using boxplot		
Unit:5	**Data Distribution**	**12 hours**
Data Distribution: Data standardizing – Z Score – Negative Z Score - Normalized Distribution– Probability Distributions - Probability of mean – location of mean distribution - Sampling Distributions — Standard Error - Standard Deviation of sampling distribution – Ratio of Sampling Distribution - Regression Analysis – Logistic Regression – Multiple Regression - ANNOVA Model – Parametric test - Non-Parametric Test		
Unit:6	**Contemporary Issues**	**2 hours**
Write an assignment on any one of the following: Analyze Global Datasets to understand Issues on Climate Change, Epidemic and Pandemic Outburst		
	Total Lecture hours	**60 hours**

Text Book(s)

1	Jiawei Han, MichelineKamber, "**Data Mining Concepts and Techniques**", Morgan Kaufmann Publishers, 2012
2	Pieter Adriaans, DolfZantinge, "**Data Mining**", Addison Wesley, 2008.
3	Krzyszlof J Cios, WitoldPedrycz, "**Data Mining: A Knowledge Discovery Approach**", Springer, 2010.
4	V. Bhuvaneswari, "**Data Analytics with R – Step by Step"**, SciTech Publications, 2016.
5	Roger D. Peng, "**R Programming for Data Science**" Lean Publishing, 2014
6	Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters,"**A Beginner's Guide to R**" Springer, 2009
7	Hadley Wickham, **"R for Data Science: Import, Tidy, Transform, Visualize, and Model Data",**First Edition, O'Reilly Media Publisher, ISBN: 9781491910399, 2017

Reference Books

1	Arun K Pujari, "**Data Mining Techniques**", Universities Press. 2012
2	ArijayChaudhry, Dr. P .S Deshpande, "**Multidimensional Data Analysis and Data Mining**", Dreamtech press, 2009.
3	Brett Lantz, **"Machine Learning with R", Third Edition,** ISBN: 9781788295864, 2019, [Packt]
4	Kaelen Medeiros, "**R Programming Fundamentals**", ISBN: 9781789612998, 2018, [Packt]
5	VitorBinanchiLanzetta, "**Hands-On Data Science with R**", ISBN: 9781789139402, 2018, [Packt]
6	Omar Trejo Navarro, "**R Programming by Example**", ISBN: 9781788292542, 2017, [Packt]
7	**Jared P. Lander, "R for Everyone: Advanced Analytics and Graphics", Second Edition, Pearson Education Publisher, ISBN: 9789386873521, 2018**
8	VigneshPrajapati, **"Big Data Analytics with R and Hadoop",** First Edition, PACKT Publishing Limited , ISBN: 9781782163282, 2013
9	Nina Zumel, **"Practical Data Science with R",**Dreamtech Press Publisher, ISBN: 9789351194378, 2014

10	Hadley Wickham, "**Advanced R**", Second Edition, CRC Publisher, ISBN 978-0815384571, 2019

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1.	Data Visualization	4 Weeks
2.	Text Retrieval and Search Engines	6 Weeks
3.	Text Mining and Analysis	6 Weeks
4.	Pattern Discovery in Data Mining	4 Weeks
5.	Cluster Analysis in Data Mining	4 Weeks
6.	Data Mining Project	6 Weeks
7.	R Programming	4 Weeks
8.	Data Analysis with R	8 Weeks
9.	Introduction to Data Analytics	9 Weeks
10.	Introduction to R Software	9 Weeks

Web Link

1. http://www.celta.paris-sorbonne.fr/anasem/papers/miscelanea/InteractiveDataMining.pdf
2. https://www.javatpoint.com/data-mining-world-wide-web
3. https://www.peterindia.net/DataMiningLinks.html
4. https://www.datacamp.com/tracks/r-programming
5. https://www.tutorialspoint.com/r/index.htm
6. https://www.datamentor.io/r-programming/

Course Designed By: Prof. Dr. V. Bhuvaneswari

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	-	M	-	M	-	M
CO2	S	S	S	S	-	S	S	M	-	M
CO3	S	S	S	S	S	M	S	L	-	L
CO4	S	S	S	S	S	S	S	S	M	S
CO5	S	S	S	S	M	M	S	M	-	S

*S-Strong; M-Medium; L-Low

Course Code		DATA VISUALIZATION	L	T	P	C
Core/Elective/Supportive		Core	4			4
Pre-requisite		Nil	Syllabus Version		2023-2024	

Course Objectives:

The main objectives of this course are to:

1. To understand how accurately represent voluminous complex data set in web and from other data sources.
2. To understand the methodologies used to visualize large data sets
3. To know how to work with visualization tools.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand the concepts of visualization	K2
2	Understand the methods for visualizing data in D3j, c3j, and Tableau	K1, K2
3	Apply Visualization methods for different data domains	K4
4	Design Interactive Charts based on Data	K3
5	Distinguish and Suggest the appropriate data visualization tools for domain specific applications and Design an Interactive data visualization story board for data	K4, K5

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** - Create

Unit:1	Introduction to Data Visualization	12 hours

Definition – Methodology – Seven Stages of Data Visualization - Data Visualization Tools. Visualizing Data: Mapping Data onto Aesthetics – Visualizing Amounts - Visualizing Distributions: Histograms and Density Plots – Visualizing Propositions: – Visualizing Associations: Among Two or More Quantitative Variables – Visualizing Time Series and Other Functions of an Independent Variable – Trends – Visualizing Geospatial Data.

Unit:2	Interactive Data Visualization	12 hours

Introduction to D3 - Fundamental Technology: The Web – HTML – DOM – CSS – JavaScript – SVG. D3 Setup – Generating Page Elements – Binding Data - Drawing with data – Scales: Domains and Ranges – Normalization – Creating a Scale – Scaling the Scatter Plot – Other Methods and Other Scales. Axes – Modernizing the Chart – Update the Data – Transition – Updates – Interactivity.

Unit:3	D3 Based Reusable Chart Library	12 hours

Setup and Deployment – Generate Chart – Customize Chart: Additional Axis – Show Axis Label – Change Chart Type – Format Values – Size – Color – Padding –Tooltip. Use APIs: Load and Unload – Show and Hide – Focus – Transform – Groups – Grid – Regions – Flow – Revert – Toggle –Legend – Sub chart – Zoom – Resize. Customize Style. Building Real time and Live Updating animated graphs with C3.

Unit:4	Data Visualization Tools : Tableau	12 hours

Environment Setup – Navigation – File & Data Types. TA SOURCE: Custom Data View – Extracting Data – Fields Operations – Editing Meta Data – Data Joining – Data Blending. Worksheets.- Bar Chart – Line Chart – Pie Chart – Scatter Plot – Bubble Chart –Gantt Chart – Histograms - Waterfall Charts. Dashboard – Formatting – Forecasting – Trend Lines – Creating Dashboard

Unit:5	Power BI	10 hours

Power BI Features – Data Slicers – Data Transformation- Field Aggregation – Transformation before Load – Formatting Currecy –transforming Columns – Mapping map with GIS – Filtering – Visualizing – Creating Dashboard (Specific Usecase or Dataset) – Publishing to Web		
Unit:6	**Contemporary Issues**	**2 hours**
Apply Visualization methods for different domains. Expert lectures, online seminars – webinars		
	Total Lecture hours	**60 hours**

SCAA DATED-18.05.2023

Text Book(s)

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.
4	Claus O. Wilke, **"Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures"**, O.Reilly, 2019.
5	Dan Clark, Beginning Microsoft Power BI: A Practical Guide to Self-Service Data Analytics , 22 February 2020

Reference Books :EBooks

1	Ritchie S. King, **"Visual Storytelling with D3: An Introduction to Data Visualization in JavaScript"**, Addison-wesley Data and Analytics, 2014.
2	Elijah Meeks, **"D3.js in Action: Data visualization with JavaScript"**, Second Edition, Manning Publications, 2017.
3	Lindy Ryan, **"Visual Data Storytelling with Tableau"**, 1st Edition, Pearson, 2018.

	Course Title	Duration	Provider
1.	Fundamentals of Visualization with Tableau	4 Weeks	Coursera

Web link

1. https://c3js.org/gettingstarted.html
2. https://www.tutorialspoint.com/tableau/index.htm
3. https://www.dashingd3js.com/table-of-contents
4. https://www.udacity.com-Data Visualization and D3.J
5. https://data-flair.training/blogs/power-bi-tutorial/

Course Designed by: Dr. S. Gavaskar

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	M	M	M	S	L	S	S	M	S
CO3	S	M	M	M	S	L	S	S	M	S
CO3	S	S	M	S	S	M	S	S	M	S
CO4	S	S	S	S	S	M	S	S	M	S
CO5	S	S	M	S	S	M	S	S	M	S

*S-Strong; M-Medium; L-Low

Course code		PRACTICAL III : DATA ANALYTICS WITH R LAB	L	T	P	C

Core/Elective/Supportive	Core	5	4
Pre-requisite	Basics of statistical Concepts and Algebra	**Syllabus Version**	2023-2024

Course Objectives:

1. To enable the students to learn the visualization package R Tool.
2. To understand & write programs using the DM algorithms
3. To apply statistical interpretations for the solutions
4. Able to use visualizations techniques for interpretations

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Able to write programs using R Tool.	K1,K2
2	To implement Packages and functions in R	K2, K3
3	Able to use different visualizations techniques using R	K3,K4
4	To apply different data mining algorithms to solve real world applications	K5 ,K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

LIST OF PROGRAMS	**75 hours**

1. Programs using basic structures of R-Data Frames, Matrix,List.
2. Programs using Packages and Functions in R
3. Exercises to gather the insights and inference from a dataset.
4. Programs using statistical functions in R
5. Programs to implement regression using R.
6. Implement Apriori algorithm to extract association rule of datamining.
7. Data Visualization

Expert lectures, online seminars – webinars

Total Lecture hours	**75 hours**

Text Book(s)

1	V. Bhuvaneswari, "**Data Analytics with R – Step by Step**", SciTech Publications, 2016.
2	Roger D. Peng, "**R Programming for Data Science**" Lean Publishing, 2014
3	Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters,"**A Beginner's Guide to R**" Springer, 2009
4	Hadley Wickham, **"R for Data Science: Import, Tidy, Transform, Visualize, and Model Data",** First Edition, O'Reilly Media Publisher, ISBN: 9781491910399, 2017

Reference Books:

1	Brett Lantz, **"Machine Learning with R", Third Edition,** ISBN: 9781788295864, 2019, [Packt]
2	Kaelen Medeiros, "**R Programming Fundamentals**", ISBN: 9781789612998, 2018, [Packt]
3	VitorBinanchiLanzetta, "**Hands-On Data Science with R**", ISBN: 9781789139402, 2018, [Packt]
4	Omar Trejo Navarro, "**R Programming by Example**", ISBN: 9781788292542, 2017, [Packt]

5	**Jared P. Lander**, "R for Everyone: Advanced Analytics and Graphics", Second Edition, Pearson Education Publisher, ISBN: 9789386873521, 2018
6	VigneshPrajapati, **"Big Data Analytics with R and Hadoop",** First Edition, PACKT Publishing Limited , ISBN: 9781782163282, 2013
7	Nina Zumel, **"Practical Data Science with R",** Dreamtech Press Publisher, ISBN: 9789351194378, 2014
8	Hadley Wickham, **"Advanced R"**, Second Edition, CRC Publisher, ISBN: 978-0815384571, 2019

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

S. No	Course Title	Duration	Provider -Free
1.	R Programming	4 Weeks	Coursera
2.	Data Analysis with R	8 Weeks	Udacity
3.	Introduction to Data Analytics	9 Weeks	Swayam
4.	Introduction to R Software	9 Weeks	Swayam
5.	Data Science Certification Training – R Programming	14 hours	Simlilearn

Web Link:

1. https://www.datacamp.com/tracks/r-programming
2. https://www.tutorialspoint.com/r/index.htm
3. https://www.datamentor.io/r-programming/

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	S	S	S	M	-	S
CO2	S	M	M	S	S	S	-	-	-	-
CO3	S	S	S	S	M	M	-	L	-	-
CO4	S	S	S	S	S	S	S	S	-	S

*S-Strong; M-Medium; L-Low

Course code		PRACTICAL IV : DATA VISUALIZATION LAB	L	T	P	C

Core/Elective/Supportive	Core	5	4
Pre-requisite	Basic Programming language	**Syllabus Version**	2023-2024

Course Objectives:

The main objectives of this course are to:

1. To understand how accurately represent voluminous complex data set in web and from other data sources.
2. To understand the methodologies used to visualize large data sets
3. To know how to work with visualization tools.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Applying methods for visualizing data in D3j, c3j, and Tableau, Power BI	K3, k4
2	Apply Visualization methods for different data domains	K3,K4
3	Design Interactive Charts based on Data	K2,K3
4	Apply the appropriate data visualization tools for domain specific applications and Design an Interactive data visualization story board for data	K4, K5,K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

LIST OF PROGRAMS	**75 hours**

1. Visualize prediction related analysis using visualization tool
2. Design a Sales forecast analysis dashboard
3. Analyze the dataset of marketing campaigns and visualize the performance
4. Analyze the product related information
5. Analyze the dataset of various crimes
6. Demonstrate 3D plotting
7. Demonstrate scatter plotting
8. Visualize business intelligence

Total Lecture hours	**75 hours**

Text Book(s)

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.
4	Claus O. Wilke, **"Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures"**, O.Reilly, 2019.

Reference Books : EBooks

1	Ritchie S. King, **"Visual Storytelling with D3: An Introduction to Data Visualization in JavaScript"**, Addison-wesley Data and Analytics, 2014.
2	Elijah Meeks, **"D3.js in Action: Data visualization with JavaScript"**, Second Edition, Manning Publications, 2017.
3	Lindy Ryan, **"Visual Data Storytelling with Tableau"**, 1st Edition, Pearson, 2018.

	Course Title	Duration	Provider
1.	Fundamentals of Visualization with Tableau	4 Weeks	Coursera
Web link			

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

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	M	M	M	S	L	S	S	-	S
CO2	S	S	M	S	S	M	S	S	-	S
CO3	S	S	S	S	S	M	S	S	-	S
CO4	S	S	M	S	S	M	S	S	-	S

*S-Strong; M-Medium; L-Low

		BIG DATA FRAMEWORKS AND TOOLS	L	T	P	C
Core/Elective/Supportive		Core	**4**			**4**
Pre-requisite		**Basics of Programming**	**Syllabus Version**		**2023-2024**	

Course Objectives: SCAA DATED: 18.05.2023

1. To understand MapReduce programming architecture, processing models.
2. To understand and design MapReduce Programming using PIG and Hive
3. To understand and compare the architectural and processing of MapReduce Programming languages Pig, Hive and SPARK

Expected Course Outcomes:

1	Understand distributed, MapReduce Processing architectures	K2
2	Configure and setup MapReduce Processing architectures Ecosystem – Hadoop, Spark , Pig and Hive	K1, K2
3	Understand and write MapReduce program using Pig and Hive, SPARK	K3
4	Critically Analyze dataset using Pig , Hive and SPARK and suggest MapReduce Programming models based on domains specific applications	K3
5	Design and setup a Big Data Analytics Ecosystem for specific Business scenarios.	K4 , K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	**Big Data Framework**	**12-- hours**

Introduction to Big Data – Distributed file system –,Hadoop Storage [HDFS], Common Hadoop Shell commands - Anatomy of File Write and Read, NameNode, Secondary NameNode, and DataNode - Map Reduce Architecture -Hadoop Configuration: Environment : Steps – Hadoop 1.0 Version VsHadoop 2.0 YARN – Setting up Hadoop Eco System – Oozie – FLUME- STORM – FLUME - Pig Configuration – Hive Configuration - SPARK Configuration – Integration – Hadoop with R – Hadoop with Python

Unit:2	**PIG : MapReduce**	**12-- hours**

Pig Introduction: Overview of Pig - Pig Architecture - Pig Execution modes, Pig Grunt shell and Shell - commands. Pig Latin Basis: Data model, Data Types, Operator - Pig Latin Commands - Load & Store, Diagnostic Operators, Grouping, Cogroup, Joining, Filtering, Sorting, Splitting - Built-In Functions, User define functions.- Pig Execution Modes – Batch Mode – Embedded Mode – Pig Execution in Batch Mode – Embedding Pig in Python – Use cases - Map Reduce programs with Pig – Pig Vs SQL

Unit:3	**Hive: Map Reduce - CURD**	**10-- hours**

Introduction of Hive - Hive Features - Hive architecture -Hive Meta store - Hive data types – Hive Tables - Table types - Creating database , Altering database, Create table, alter table, Drop table, - Built-In Functions - Built-In Operators, User defined functions –

Unit:4	**Hive: Aggregation and Indexing**	**12-- hours**

HiveQL–Introduction to HiveQL, HiveQL Select, HiveQL – MapReduce using HiveQLOrderBy,Group By Joins, LIMIT, Distribute By , Cluster By - Sorting And Aggregation – Partitioning – Static –Dynamic – Index Creation - Bucketing – Analysis of MapReduce execution – Hive Optimization – Setting Hiving Parameters. – Usecase :MapReduce using Hive QL – HiveQLVs SQL

Unit:5	**SPARK Query**	**12-- hours**

SPARK – MapReduce - RDD Transformations – SPARK Operations – Usecase with SPARK and Comparison - MapReduce – Python – R – Pig – Spark – Hadoop - Limitations – Advantage – SPARK vsHadoop – SPARK Vs Pig and Hive – MapReduce- Spark Transformations

Unit:6	**Contemporary Issues**	**2 hours**

Expert lectures, online seminars – webinars

	Total Lecture hours	**60-- hours**

Text Book(s):

1	Boris Lublinsky Kevin T. Smith Alexey Yakubovich,ProfessionalHadoop® Solutions, Wiley, ISBN: 9788126551071,2015.
2	Chris Eaton, Dirk deroos et al., "**Understanding Big data**", McGraw Hill, 2012.

3	Tom White, "**Hadoop: The Definitive Guide**", O'Reilly Media 3rd Edition, May 6, 2012
4	Donald Miner, Adam Shook, "**MapReduce Design Patterns**", O'Reilly Media November 22, 2012
5	Edward Capriolo,DeanWampler,Jason Rutherglen, "**Programming Hive**", O'Reilly Media; 1 edition , October, 2012
6	Deepak Vohra, "Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools" First Edition, Apress Publisher, ISBN: 9781484221983, 2016
7	Alan Gates, "**Programming Pig**", O'Reilly Media; 1st Edition,October, 2011
Reference Books:	
1	Sridhar Alla**, "Big Data Analytics with Hadoop 3",** First Edition, ISBN: 978-1-78862-884-6, 2018, [Packt]
2	Naresh Kumar, **"Modern Big Data Processing with Hadoop",** ISBN: 9781787122765, 2018, [Packt]
3	NeerajMalhotra, **"Data Engineering Skills - Hadoop Shell: A Comprehensive Guide to Hadoop FS Commands",** First Edition, CreateSpace Independent Publishing, ISBN: 9781717577511, 2018
4	VigneshPrajapati, **"Big Data Analytics with R and Hadoop",** First Edition, ISBN: 978-1-78216-328-2, 2013, [Packt]
5	Edward Capriolo, "**Programming Hive: Data Warehouse and Query Language for Hadoop",** First Edition, O'Reilly MediaPublisher, ISBN**:** 9781449319335, 2012

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

S. No	Course Title	Duration	Provider - Free
1.	Big Data Hadoop and Spark Developer – R Programming	26 hours	Simplilearn
2.	Intro to Hadoop and MapReduce	4 Weeks	Udacity
3.	Hadoop Platform and Application Framework	5 Weeks	Coursera
4.	Big Data Essentials: HDFS, MapReduce and Spark RDD	6 Weeks	Coursera
5.	Mining Massive Datasets	7 Weeks	edX

Web Link – Video

1. http://hadooptutorial.info/mapreduce-programming-model/
2. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
3. https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html
4. https://www.edureka.co/blog/mapreduce-tutorial/

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	-	M	M	-	-	-	-	M	-	-
CO2	S	S	S	S	M	S	-	M	M	M
CO3	M	M	M	S	-	-	-	M	-	M
CO4	S	S	S	S	M	M	S	L	M	S
CO5	S	S	S	S	S	S	S	S	S	S

*S-Strong; M-Medium; L-Low

		MACHINE LEARNING	L	T	P	C
Core/Elective/Supportive		Core	**4**			**4**

Pre-requisite	Basics on Statistics and Linear Algebra	Syllabus Version	2023 - 2024

Course Objectives:

1. To understand the Concepts of Machine learning algorithms.
2. To apply the machine learning algorithms for various applications.

Expected Course Outcomes:

CO1	Understand the concepts of machine learning.	K1
CO2	Understand the various supervised learning techniques	K2
CO3	Understand the theoretical concepts of linear methods	K2
CO4	Apply Supervised, Unsupervised and Semi Supervised learning algorithm	K4
CO5	Understand and apply the concept of Deep Learning	K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6**– Create

Unit:1	**Machine Learning**	**12-- hours**

Introduction: Machine learning – Examples and Applications - Perspectives and Issues in Machine learning - Input: Concepts, Instances, and Attributes - Output: Knowledge Representation-Credibility: Evaluating What's Been Learned: Training and Testing - Predicting Performance - Cross Validation - Other Estimates - Counting the cost.

Unit:2	**Decision Tree**	**12-- hours**

Decision Tree Learning: Decision tree representation – Decision tree learning – Random forest -Issues in decision tree learning- Bayesian Learning: Naïve Bayes classifier - Instance Based Learning: Introduction – k- nearest neighbor Learning - Radial Basis Function, Case based reasoning.

Unit:3	**Artificial Neural Networks**	**12-- hours**

Artificial Neural Network - Introduction – Neural Network Representation - Perceptrons – Multilayer Networks and Backpropagation Algorithm - Linear models for Regression- Linear Discriminant Analysis - PCA – Kernel PCA

Unit:4	**SVM**	**12-- hours**

SVM : Introduction – Kernel methods - formulation and computation- SVM Linear classifier – SVM with two variables –non-linear classifier-Polynomial kernels- Radial Basis Function Kernels - Clustering Methods.- Introduction – K- Means- Expectation-Maximization Algorithm- Hierarchical Clustering - Choosing the Number of Clusters.

Unit:5	**Deep Learning**	**10-- hours**

Deep Learning – Deep feed forward network – Convolutional neural network– Autoencoders –Deep Belief Networks -Recurrent Neural Network - Use Cases: Finding similar users in Twitter (Mahout), Email marketing system (Mahout).

Unit:6	**Contemporary Issues**	**2 hours**

Expert lectures, online seminars – webinars

	Total Lecture hours	**60-- hours**

Text Books:

1	Ian Witten, Data mining: Practical Machine Learning Tools and Techniques, Fourth edition, Morgan Kaufmann Publishers.
2	Tom M. Mitchell (1997). Machine Learning, Tata McGraw-Hill, New Delhi

3	K.P. Soman, Machine Learning with SVM and Other Kernel Methods, 2011, PHI Publishing
Reference Books	
1	JannesKlaas, "Machine Learning for Finance", ISBN: 978178936364, 2019 [Packt]
2	Giuseppe Bonaccorso, "Machine Learning Algorithms", Second Edition, ISBN: 9781789347999, 2018 [Packt]
3	Stephen Marsland, "Machine Learning –An Algorithmic Perspective", CRC Press, 2009
4	Hastie, Tibshirani, Friedman, "The Elements of Statistical Learning", Second Edition, Springer, 2008
5	Yuxi Liu, "Python Machine Learning By Example", 2017 [Packt]
6	John Paul Mueller, Luca Massaron, "**Machine Learning (in Python and R) For Dummies", First Edition,** Wiley Publisher, ISBN: 9788126563050, 2016
7	U Dinesh Kumar Manaranjan Pradhan**,,"Machine Learning using Python".**) Publisher: Wiley, ISBN: 9788126579907, 2019

Online Course:			
S. No	**Course Title**	**Duration**	**Provider -Free**
1.	Machine Learning	12 hours	Simplilearn
2.	Machine Learning for Data Analysis	4 Weeks	Coursera
3.	Machine Learning Foundations: A Case Study Approach	6 Weeks	Coursera
4.	Machine Learning : Regression	6 Weeks	Coursera
5.	Introduction to Machine Learning	12 Weeks	Swayam - NPTEL
6	Deep Learning Specialization	4 Courses	Coursera

Web Link - Video:

1. https://www.packtpub.com/data/hands-on-machine-learning-with-scikit-learn-and-tensorflow-2-0-video
2. https://www.packtpub.com/data/machine-learning-projects-with-tensorflow-2-0-video
3.https://www.packtpub.com/application-development/complete-machine-learning-course-python-video

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	L	L	M	M	M	L
CO2	S	S	S	-	-	-	-	-	-	L
CO3	S	S	-	-	-	-	L	L	L	L
CO4	S	S	S	S	L	M	M	M	M	M
CO5	S	S	S	S	S	S	S	S	S	S

*S-Strong; M-Medium; L-Low

		BUSINESS ANALYTICS	**L**	**T**	**P**	**C**
Core/Elective/Supportive		**Core**	**4**			**4**
Pre-requisite		**Foundations of Data Science**	**Syllabus Version**		**2023-2024**	

Course Objectives: SCAA DATED: 18.05.2023

The main objectives of this course are to:

1. To understand the Probability Theory
2. To understand theoretical distributions and automata theory

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Compare various domain areas and their challenges	K2
2	Apply the concepts of analytics to make better decisions	K3
3	Examine use cases for different domains.	K4
4	Evaluate the challenges faced in various domains and choose appropriate analytics solutions in all domains	K5
5	Propose suitable analytics solutions as required by the use cases.	K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** - Create

Unit:1		**12-- hours**

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

Unit:2		**12-- hours**

Banking and Finance: Systems of Banking – Commercial Banking – New Financial Services:Overview of Analytics in Insurance: Key Insurance Analytics – Challenges – Health Insurance Analytics, Life Insurance Analytics- Types of Insurance – Housing Finance.

Unit:3		**10 -- hours**

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

Unit:4		**12-- hours**

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

Unit:5		**12-- hours**

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

Unit:6	**Contemporary Issues**	**2 --hours**

Expert lectures, online seminars – webinars

	Total Lecture hours	**60-- hours**

Reference Books

1	Dwight McNeill(2013). A Framework for Applying Analytics in Healthcare: What Can Be Learned from Best Practices in Banking , Retail, Politics and Sports, Pearson Education
2	Gomez Clifford(2011). Banking and Finance Theory Law and practice, PHI Learning
3	Patricia L.Saporito(2014). Applied Insurance Analytics: A Framework for Driving More Value from Data Assets, Technologies and Tools, Pearson Education LTD
4	Anders Olsson(2005). Understanding Changing Telecommunications, Wiley Publications
5	Jennifer LeClaire, Danielle Dahlstrom, Vivian Braun. Business analytics in Retail for dummies, 2nd IBM Limited edition.
6	Purba Halady Rao (2013). Business Analytics. An application Focus, PHI Learning private ltd.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

Mapping with Programme Outcomes										
Cos	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**
CO1	S	M	M	L	-	-	-	-	-	M
CO2	S	M	S	S	M	-	S	L	-	M
CO3	S	S	S	L	-	L	M	L	-	L
CO4	M	M	S	S	-	-	S	-	-	L
CO5	S	S	S	M	-	L	M	L	-	L

***S-Strong; M-Medium; L-Low**

Course code		**PRACTICAL V : BIG DATA ANALYTICS LAB**	**L**	**T**	**P**	**C**
Core/Elective/Supportive		**Core**			5	4
Pre-requisite		**Basics of Programming**	**Syllabus Version**		**2023-2024**	

Course Objectives: SCAA DATED: 18.05.2023

1. To understand MapReduce programming architecture, processing models.
2. To understand and design MapReduce Programming using PIG and Hive
3. To understand and compare the architectural and processing of MapReduce Programming languages Pig, Hive and SPARK

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Configure and setup MapReduce Processing architectures Ecosystem – Hadoop, Spark, Pig and Hive	K1, K2
2	Create a MapReduce program using Pig and analyse dataset using Pig Latin Scripts	K3,K4,K5,K6
3	Apply Hive commands on a dataset	K3
4	Develop a MapReduce program using SPARK	K3,K4 , K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

LIST OF PROGRAMS	**75 hours**

1. Installation of Hadoop, Spark, Pig and Hive
2. File Management tasks in Hadoop
3. Word Count Map Reduce program to understand Map Reduce Paradigm
4. Pig Latin scripts to sort, group, join, project, and filter your data.
5. Hive Databases, Tables and Views
6. Hive Functions and Indexes
7. Hive UDFs (User Defined Functions)
8. Exercises on SPARK MapReduce

Expert lectures, online seminars – webinars

Total Lecture hours	**75 hours**

Text Book(s):

1	Boris Lublinsky Kevin T. Smith Alexey Yakubovich,ProfessionalHadoop® Solutions, Wiley, ISBN: 9788126551071,2015.
2	Chris Eaton, Dirk deroos et al., "**Understanding Big data**", McGraw Hill, 2012.
3	Tom White, "**Hadoop: The Definitive Guide**", O'Reilly Media 3rd Edition,May 6, 2012
4	Donald Miner, Adam Shook, "**MapReduce Design Patterns**", O'Reilly Media November 22, 2012
5	Edward Capriolo,DeanWampler,Jason Rutherglen, "**Programming Hive**", O'Reilly Media; 1 edition , October, 2012
6	**Deepak Vohra**, "Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools" First **Edition,** Apress Publisher, ISBN: 9781484221983, 2016
7	Alan Gates, "**Programming Pig**", O'Reilly Media; 1st Edition,October, 2011

Reference Books:

1	Sridhar Alla**, "Big Data Analytics with Hadoop 3",** First Edition, ISBN: 978-1-78862-884-6, 2018, [Packt]

2	Naresh Kumar, **"Modern Big Data Processing with Hadoop",** ISBN: 9781787121768, 2018, [Packt]
3	NeerajMalhotra, **"Data Engineering Skills - Hadoop Shell: A Comprehensive Guide to Hadoop FS Commands",** First Edition, CreateSpace Independent Publishing, ISBN: 9781717577511, 2018
4	VigneshPrajapati, **"Big Data Analytics with R and Hadoop",** First Edition, ISBN: 978-1-78216-328-2, 2013, [Packt]
5	Edward Capriolo, "**Programming Hive: Data Warehouse and Query Language for Hadoop",** First Edition, O'Reilly MediaPublisher, ISBN**:** 9781449319335, 2012

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

S. No	Course Title	Duration	Provider - Free
1.	Big Data Hadoop and Spark Developer – R Programming	26 hours	Simplilearn
2.	Intro to Hadoop and MapReduce	4 Weeks	Udacity
3.	Hadoop Platform and Application Framework	5 Weeks	Coursera
4.	Big Data Essentials: HDFS, MapReduce and Spark RDD	6 Weeks	Coursera
5.	Mining Massive Datasets	7 Weeks	edX

Web Link – Video

1. http://hadooptutorial.info/mapreduce-programming-model/
2. https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
3. https://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html
4. https://www.edureka.co/blog/mapreduce-tutorial/

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	S	M	S	-	M	-	M
CO2	M	M	M	S	-	-	-	M	-	M
CO3	S	S	S	S	M	M	S	L	-	S
CO4	S	S	S	S	S	S	S	S	-	S

*S-Strong; M-Medium; L-Low

Course code		PRACTICAL VI : MACHINE LEARNING LAB	L	T	P	C
Core/Elective/Supportive		Core			5	4
Pre-requisite		Basics of Machine Learning	Syllabus Version		2023-2024	

Course Objectives:

The main objectives of this course are to:
1. Build models using classification algorithm for real world problems
2.Build models using clustering algorithm for real world problems
3. Create classification and clustering models
4. Test and evaluate the models

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand the various supervised learning techniques	K2
2	Understand the theoretical concepts of linear methods	K2
3	Apply Supervised, Unsupervised and Semi Supervised learning algorithm	K4
4	Understand and apply the concept of Deep Learning	K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** - Create

LIST OF PROGRAMS	75 hours

Implement the following in Python:

1.Exercise to manipulate data using different queries
2.Exercises to load dataset into sci-kit learn
3.Exercise for Building models in sci-kit learn
4.Exercise to extract features from datasets
5.Exercise to implement Regression
6.Exercise to implement SVM Classifier
7.Exercise to implement K-Means Clustering
8.Exercises for Deep learning
9.Exercises to Build a data pipeline

Total Lecture hours	75 hours

ELECTIVES

		TEXT ANALYTICS	L	T	P	C
Core/Elective/Supportive		Elective	**4**			**4**

Pre-requisite		Syllabus Version	2023-2024

Course Objectives:

1. To understand the text mining and NLP techniques.
2. To understand and applyprobabilistic models, clustering and classification for text analytics.
3. To understand and apply text analytics approaches in different domains.
4. To understand representation and handling of opinions by people in different ways.
5. To analyse different challenges in sentiment analysis and aspect-oriented sentiment analysis classification and analyse fake opinion detection and intention classification

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand the concepts of text mining and text pre-processing techniques	K1, K2
2	Apply the probabilistic models, clustering and classification for text analytics	K3
3	Design a text analytic framework to analyze text data for domain specific applications	K4, K5 K6
4	Introduction to sentiment analysis and its applications	K1,K2
5	Create different types of opinion summary from the given data sources	K1,K3
6	Identifying opinion quality, author intention and fake opinions	K1,K4

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6**– Create

Unit:1	**Text Mining**	**10-- hours**

Text Mining - Definition - General Architecture – Core Text mining Operations. Nature of unstructured and semi-structured text, collecting documents NLP: Text pre-processing-Sentence Segmentation tokenization - lemmatization - stemming - Parsing text - keywords- POS, Bag of Words Model, n-grams, chunking and Named Entity Recognition (NER) Corpus - sentence boundary determination - Textual information to numerical vectors -vector generation for prediction- document standardization and Representation – Inverted Index-term document matrix (TDM)-TDM Frequency

Unit:2	**Information retrieval and Extraction**	**12-- hours**

Information retrieval- keyword search - Vector space scoring, Models - web- based document search-matching-inverted lists. Information extraction-Architecture - Co-reference - Named Entity and Relation Extraction-Template filling and database construction –Applications. Inductive -Unsupervised Algorithms for Information Extraction.

Text Categorization – Definition – knowledge engineering,Text Classification, Feature Selection for Text Classification, Gini Index, Information Gain. Evaluating model: confusion matrix, class specific measure Classification models: Decision Tree Classifiers -Rule- based Classifiers - Naive Bayes Classifiers - Methods for Text Clustering –Distance and similarities

.

Unit:3	**Probabilistic Models for Text Mining**	**12-- hours**

Probabilistic Models: Introduction, Mixture Models, Stochastic Processes in Bayesian Nonparametric Models, Graphical Models, Probabilistic Models with Constraints, Parallel Learning Algorithms. Probabilistic Models for Information Extraction -Hidden Markov Models -Stochastic Context-Free Grammars - Maximal Entropy Modeling -Maximal Entropy Markov Models - Conditional Random Fields

Unit:4	Sentiment Analysis	12-- hours

Introduction: Sentiment Analysis Applications - Sentiment Analysis Research - Sentiment Analysis as Mini NLP. The Problem of Sentiment Analysis: Definition of Opinion - Definition of Opinion Summary - Affect, Emotion, and Mood - Different Types of Opinions - Author and Reader Standpoint. Document Sentiment Classification: Supervised Sentiment Classification - Unsupervised Sentiment Classification - Sentiment Rating Prediction - Cross-Domain Sentiment Classification - Cross-Language Sentiment Classification - Emotion Classification of Documents.

Unit:5	Subjectivity Classification and Challenges	12-- hours

Subjectivity - Sentence Subjectivity Classification - Sentence Sentiment Classification - Dealing with Conditional Sentences - Dealing with Sarcastic Sentences - Cross-Language Subjectivity and Sentiment Classification - Using Discourse Information for Sentiment Classification - Emotion Classification of Sentences. Subjectivity classification and Aspect Based sentiment classification. Sentiment Lexicon Generation: Dictionary-Based Approach - Corpus-Based Approach - Desirable and Undesirable Facts. -

Use Cases: Detecting Fake or Deceptive Opinions: Different Types of Spam - Supervised Fake Review Detection - Supervised Yelp Data Experiment - Automated Discovery of Abnormal Patterns - Model-Based Behavioral Analysis - Group Spam Detection - Identifying Reviewers with Multiple User ids - Exploiting Business in Reviews - Some Future Research Directions.

Unit:6	Contemporary Issues	2-- hours

Challenges of text analytics approaches for regional specific languages

	Total Lecture hours	60-- hours

Text Book(s)

1	Murugan Anandarajan "Practical Text Analytics: Maximizing the Value of Text Data", Springer; 2018
2	Charu C. Aggarwal Machine Learning for Text 2018
3	Steven Bird, Ewan Klein and Edward Loper" Natural Language Processing with Python"
4	Bing Liu "Sentiment Analysis: Mining Opinions, Sentiments and Emotions, Cambridge University Press, 2015.

Reference Books

1	Markus Hofmann, Andrew Chisholm "Text Mining and Visualization: Case Studies Using Open-Source Tools,", CRC press, Taylor & Francis,2016
2	Charu C. Aggarwal ,Cheng Xiang Zhai, Mining Text Data, Springer; 2012
3	Dipanjan Sarkar Text Analytics with Python, 2016
4	Bing Liu "Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.
5	Erik Cambria, Dipankar Das "A Practical Guide to Sentiment Analysis" Springer, 2017.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1	Business Analytics & Text Mining Modelling Using Python, IIT Roorkeehttps://swayam.gov.in/
2	Natural Language Processing, IIT Kharagpurhttps://swayam.gov.in/
3	Text Mining and Natural Language Processing in Rhttps://www.udemy.com/

Course Designed By:

Mapping with Programme Outcomes										
COs	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**
CO1	M	M	M	-	-	-	S	S	S	S
CO2	M	M	M	S	-	-	S	S	S	S
CO3	S	S	S	S	S	M	S	S	S	S
CO4	S	S	S	S	S	-	S	S	S	S
CO5	S	S	S	S	S	S	S	S	S	S

*S-Strong; M-Medium; L-Low

		HEALTH CARE DATA ANALYTICS	**L**	**T**	**P**	**C**
Core/Elective/Supportive		**Elective**	**4**			**4**
Pre-requisite		**Basics on Statistics and Linear Algebra**	**Syllabus Version**		**2023-2024**	

Course Objectives:

1. To understand the Process ,Concepts and Procedures in Health Care Data Digital Systems
2. Understand Data standards used in Health Care Domain
3. Design Integrated Health Care Data Models for Data Analytics
4. Understand and Remember the Ethics of Managing and Analyzing Health Care Data

Expected Course Outcomes:

CO1	Understand the Process and Data Functionalities of Health Care Data	K1, K2
CO2	Understand the various Data Sources, diagnostic standards and Components of Data Analytics	K2, K1
CO3	Understand and design Integrated Data Model for analytics	K2, K5

CO4	Apply ETL for data analysis and create dashboards	K3, K4
CO5	Create and evaluate prediction models in healthcare applications for preventive care and personalized medicines	K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	**Health Care Systems**	**12-- hours**

Introduction :Health Care Entities – Electronic Health Care Records – Clinical Data - Health Care Big Data Sources– Patient Data – Administrative Data – Genomics Data – Imaging Data- Insurance Data – Diagnostic Data – Clinical Data— Social Media – Survey Data – Family Data – Data Quality – Data Ethics – Data Integration Challenges

Unit:2	**Data Models and Data Standards**	**12-- hours**

Data Models : Relational Models – Hierarchical Models — Data warehousing Models – Star Schema – Normalized Data and Deformalized – Health Care Knowledge Representation Ontologies – Diagnosis Standards – ICD 9/10 - DSMI – DSM II –Drug Standards SNOWMED –LOINC – Laboratory Standards – Data Challenges in Data Mapping -Data Standards as Linked Data

Unit:3	**Big Data and Data Analytics**	**10-- hours**

Data Analytics: Data Cleaning and Pre-Processing – Data Processing and Modeling - Classification – Clustering – Dimensionality Reduction - Prediction Machine Learning – Microsoft Azure Cloud -Data Visualizing – Histogram – Boxplot- Scatter Plot – Bar – Pie – Mosaic Plot – Trends Lines – Heat Maps – Density Plots - Dashboard – Creation - Presentation

Unit:4	**Advanced Health Care Analytics**	**12-- hours**

Genomics Data Analysis – Microarray Data – Sequence Data – Research Survey Analysis – Text Mining – Tele Health – Virtual Health Care Assistance

Unit:5	**Health Care Usecase**	**12-- hours**

Prediction of Risk of Co morbidity Individuals – Outbreak – Epidemics - Personalized Medical Care – Pharmaceuticals and Patient Data Integration – Clinical Data

Unit:6	**Contemporary Issues**	**2 hours**

Expert lectures, online seminars – webinars

	Total Lecture hours	**60-- hours**

Text Books:	
1	Christopher Bishop, "Pattern Recognition and Machine Learning" Springer, 2006
2	Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012
3	Ethem Alpaydin, "Introduction to Machine Learning 3(Adaptive Computation and Machine Learning Series)", Third Edition, MIT Press, 2014
4	Tom M Mitchell, "Machine Learning", First Edition, McGraw Hill Education, 2013.
Reference Books	
1	Jannes Klaas, "Machine Learning for Finance", ISBN: 978178936364, 2019 [Packt]
2	Giuseppe Bonaccorso, "Machine Learning Algorithms", Second Edition, ISBN: 2018 [Packt]
3	Stephen Marsland, "Machine Learning –An Algorithmic Perspective", CRC Press, 2009
4	Hastie, Tibshirani, Friedman, "The Elements of Statistical Learning", Second Edition, Springer, 2008
5	Yuxi Liu, "Python Machine Learning By Example", 2017 [Packt]
6	John Paul Mueller, Luca Massaron, "Machine Learning (in Python and R) For Dummies", First Edition, Wiley Publisher, ISBN: 9788126563050, 2016
7	U Dinesh Kumar ManaranjanPradhan,,"Machine Learning using Python".) Publisher: Wiley, ISBN: 9788126579907, 2019

Online Course:			
S. No	**Course Title**	**Duration**	**Provider -Free**
1.	Machine Learning	12 hours	Simplilearn
2.	Machine Learning for Data Analysis	4 Weeks	Coursera
3.	Machine Learning Foundations: A Case Study Approach	6 Weeks	Coursera
4.	Machine Learning : Regression	6 Weeks	Coursera
5.	Introduction to Machine Learning	12 Weeks	Swayam -NPTEL
6	Deep Learning Specialization	4 Courses	Coursera

Web Link - Video:

1. https://www.packtpub.com/data/hands-on-machine-learning-with-scikit-learn-and-tensorflow-2-0-video
2. https://www.packtpub.com/data/machine-learning-projects-with-tensorflow-2-0-video
3.https://www.packtpub.com/application-development/complete-machine-learning-course-python-video

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	M	M	M	-	-	-	S	S	S	S
CO2	M	M	M	S	-	-	S	S	S	S
CO3	S	S	S	S	S	M	S	S	S	S
CO4	S	S	S	S	S	-	S	S	S	S
CO5	S	S	S	S	S	S	S	S	S	S

*S-Strong; M-Medium; L-Low

		BEHAVIOURAL DATA ANALYTICS	**L**	**T**	**P**	**C**
Core/Elective/Supportive		**Elective**	**4**			**4**

Pre-requisite	Basics on Statistics and Linear Algebra	Syllabus Version	2023-2024

Course Objectives:

1) To familiarize the student with issues and applications of ABA and behavioral consultation in Education and Business settings.
2) Design skill acquisition programs based on the Statistical Assessment
3) Recognize and provide examples of the elementary for verbal and nonverbal operant.

Expected Course Outcomes:

CO1	To Understand the concept of Behaviour Analytics	K1, K2
CO2	To Understand about the Concepts of verbal and non-verbal Behaviour	K2, K3
CO3	To Understand the Statistical Approaches to analyze Behaviour Patterns	K3, K4
CO4	Apply Exploratory Data Analytics to find the Behavioural patterns	K5,
CO5	Design and Develop Behavioural model using various Tools	K6, K3

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** - Create

On the successful completion of the course, student will be able to:

Unit:1	Behaviour Analytics	10-- hours

Introduction Behaviour Analytics – Behaviourism in Historical Context – Classical Conditioning- operant Conditioning - Modern Behaviourism - Personal Behaviour Change Activity - Analyse behaviour - Understanding Behavioural data- Self- Assessment: Recognize- Define – Measure- analyse- improve- control - Causal-Behavioural - Introduction Causal Diagrams (CD)- Building CD-Behaviour.

Unit:2	Verbal Behaviour Statistical Approach	12-- hours

DifferencesVerbalBehaviour: A Functional Analysis of Verbal Behaviour - Controlling Variables - Multiple Variables - The Manipulation of Verbal Behaviour - The Production of Verbal behaviour - Self-Editing - Logical and Scientific Verbal Behaviour – Thinking. Non-Verbal behaviour: Basic - Critical Listening Skills - Behavioural Activation Treatment - Brainstorm Activity SMART Goals - Re-evaluating - Maintaining Gain - Strengthening behaviour Change.

Unit:3	Statistical Approach	12-- hours

A/B Experimentation – A/B test Types -Statistical Approach - A/B testing Mistakes – Challenges - Funnel Analysis – Event Properties - Conversion Drivers - Purchase Conversion Funnel -Cart Conversion Funnel -Custom Event Funnel-Campaign Conversion Funnel - Cohort Analysis - Predictive Cohorts - Behavioral Cohorts - Feature Adoption - Improving Advertising Performance - Understanding Seasonal

Unit:4	Exploratory Data Analytics	12-- hours

Exploratory Data Analytics- data Exploration -Feature Engineering – Data Cleaning – Preprocessing – Missing values – Imputation- Smoothing - Normalization – Imbalance Classes- Sampling - Sampling Types- One-hot Encoding - Summary Statistics – Automated EDA: Data Explorer Package - Ensemble Learning

Unit:5	Vintage Analysis	2 hours
Vintage Analysis - Behavior Analysis: Organization – E-commerce - Use Cases: Health behavior Change – Dataset – Tools: Adobe's funnel analysis product -Google Analytics- Heap Mixpanel Report – Dashboard		
Unit:6	**Contemporary Issues**	**2 hours**
Expert lectures, online seminars – webinars		
	Total Lecture hours	**60-- hours**

Text Books:

1	Chase, P. N., and Smith, J. M. (1994). *Performance Analysis: Understanding Behavior in Organizations*. Morgantown, WV: Envision Development Group, Inc., Publishers.
2	Daniels, A. C., and Daniels, J. (2004*). Performance Management: Changing Behavior That DrivesOrganizational Effectiveness* (4thed.). Tucker, GA: Performance Management Publications.

Reference Books

1	Florent Buisson, "Behavioral Data Analysis with R and Python", (2021), Published by O'Reilly Media, Inc., 2021.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

S.No	Course Title	Duration	Provider
1.	Behavioral Analytics (Free)	5 Lectures	Coursera
2.	Applied Behavioural Analysis (ABA)	2 hours	Udemy
3.	Behavioral Psychology Courses: Leadership and organizational behavior	4 Weeks	edX
4.	Applied Behavior Analysis - Foundation Course	2 hours	Udemy

Course Designed by: Prof. Dr. V. Bhuvaneswari

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	S	-	-	-	M	M	M	M
CO2	M	M	M	S	-	-	S	S	S	S
CO3	S	S	S	S	M	M	S	S	S	S
CO4	S	S	S	S	M	S	M	S	S	S
CO5	S	S	S	S	S	S	S	S	S	S

*S-Strong; M-Medium; L-Low

		INTERNET OF THINGS	L	T	P	C
Core/Elective/Supportive		Elective	**4**			**4**
Pre-requisite		Basic knowledge of hardware, Programming in C	**Syllabus Version**		**2023-2024**	

Course Objectives:

The main objectives of this course are:

1. To gain insight about the architecture and enabling technologies of Internet of Things
2. To understand Arduino micro controller and IDE
3. To develop simple IoT Applications for different domains

Expected Course Outcomes:		SCAA DATED: 18.05.2023
On the successful completion of the course, student will be able to:		
CO1	To learn the importance of smart objects and smart environment	K1
CO2	To understand and use the microcontroller and various sensors	K2
CO3	To create programs using Arduino IDE and extract data	K3
CO4	To perform WiFi data communications, remote data storage in cloud, and handle the data using web applications	K3, K4
CO5	To identify potential problems and develop solutions using IOT	K5, K6
K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create		

Unit:1	**Introduction to IOT**	**10 hours**
Introduction to IOT - Enabling technologies of IOT - AI and Machine Learning - Physical and logical design of IoT - IOT Reference Architecture - IOT Functional Architecture - IoT levels and deployment templates – Application domains of IoT: Home automation – Cities – Environment – Energy – Industry – Agriculture – Transportation - Health care & Lifestyle.		

Unit:2	**Basic Electronics for IoT & Arduino IDE**	**12 hours**
Understanding basic electronic components and power elements Electric Charge, Resistance, Current and Voltage – Resistors, Capacitors, Diodes, LED, Potentiometer, circuit boards - Analog and digital circuits – Microcontrollers – Electronic Signals – A/D and D/A Conversion – Pulse Width Modulation Arduino IDE: Installation and Set-up - Programming Fundamentals with C using Arduino IDE Program Structure in C - Basic Syntax - Data Types / Variables / Constants - Operators, Conditional Statements and Loops - Using Arduino C Library functions for Serial, delay and other invoking functions.		

Unit:3	**Arduino Microcontroller and sensors**	**12 hours**
Working with Arduino: LED and Switch - Data acquisition with IOT Devices - Understanding Sensors and Devices - Understanding the Inputs from Sensors - Working with Temperature Sensors -Working with Ultrasound Sensor -Working with humidity sensor - Working with Motion Sensor - Working with IR Sensor - Working with Proximity Sensor - Working with Accelerometer and vibration sensor.		

Unit:4	Medical Sensors and Actuators	12 hours
Understanding Medical Sensors: Flow Sensor - Optical Sensor - Body Temperature Sensor - Blood Pressure Sensor -Airflow sensor (breathing) - Patient position sensor (accelerometer) - Pulse and oxygen in blood sensor (SPO2) - Galvanic skin response (GSR - sweating) sensor. Understanding the Outputs through Actuators - Activating LED Lights - Activating Relays - Activating Buzzer - Running DC Motors - Running Stepper Motors and Servo Motors.		
Unit:5	**Data Communication from IOT devices**	**12 hours**
Building and Using Communication Devices to transfer data from IOT Devices - Understanding the Communication Principles to Transfer the data from IOT Devices; Using WIFI to Transfer the data from IOT Sensor; Programming Fundamentals with Web Applications for handling Data Communication from IOT Device; Remote Communication to cloud/external application .		
Unit:6	**Contemporary Issues**	**2 hours**
Expert lectures, online seminars – webinars		
	Total Lecture hours	**60 hours**

Text Book(s)

1	ArshdeepBahga, Vijay Madisetti, ‘Internet of Things: A Hands-On Approach’, Universities Press, 2015.
2	Boris Adryan, DominikObermaier, Paul Fremantle, ‘The Technical Foundations of IoT’, Artech Houser Publishers, 2017.
3	Michael Margolis, “Arduino Cookbook” 2nd Edition, O'Reilly Media, 2012.
4	Marco Schwartz, ‘Internet of Things with ESP8266’, Packt Publishing, 2016.

Reference Books

1	Charles Platt, “Make Electronics – Learning by discovery”, O'Reilly Media, 2015.
2	Michael Miller, “The Internet of Things”, Pearson India, 2015.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1	Introduction to IOT,https://nptel.ac.in/courses/106/105/106105166/

Course Designed By: **Prof. Dr. T. Amudha**

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	-	-	-	-	-	-	-	-
CO2	-	L	S	S	-	-	S	-	-	-
CO3	-	M	L	L	-	-	S	M	-	M
CO4	-	L	S	S	-	-	M	M	-	M
CO5	-	L	M	M	-	-	-	L	-	L

*S-Strong; M-Medium; L-Low

		DEEP LEARNING	L	T	P	C
Core/Elective/Supportive		Elective	**4**			**4**

Pre-requisite	Nil	Syllabus Version	2023-2024

Course Objectives:

The main objectives of this course are to:

1. To understand the fundamental concepts of Deep Learning.
2. To understand the concepts of Deep Learning Categories.
3. To understand and apply Deep Learning concepts in real-time.

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand the structure and model of Deep Learning	K2
2	Understand the concepts of Neural Network and its type.	K3
3	Understand and create workstation models using Python/tensorflow	K4
4	Understand and apply concepts of Deep Learning and Deep generative model.	K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6**– Create

Unit:1	Introduction: Deep Learning	12 hours

Introduction to Deep Learning: Fundamentals of Deep Learning- Artificial Intelligence – Machine Learning – Learning process of neural Network - representation data - Methodology of Deep Learning - Data representation of Neural Networks – tensor operations – Gradient based optimization - Backpropagation components – Model Parameterization – Deep Learning hyperparameter – basic configuration.

Unit:2	Neural Network	10 hours

Anatomy of Neural Network – Introduction Keras - Setting up Deep Learning Workstation – Fundamentals of Machine Learning – Evaluating Machine Learning Models – Data Preprocessing – Feature Engineering – overfitting – Underfitting – Workflow of Machine Learning.

Unit:3	Classification of Neural Network	12 hours

Feedforward Networks: Multilayer Perceptron, Gradient Descent, Backpropagation, Empirical Risk Minimization, regularization, autoencoders - Regularization for Deep Learning, Optimization for Training Deep Models

Convolutional Networks: The Convolution Operation - Variants of the Basic Convolution Function - Structured Outputs - Data Types - Efficient Convolution Algorithms - Random or Unsupervised Features- LeNet, AlexNet

Recurrent Neural Networks: Bidirectional RNNs - Deep Recurrent Networks Recursive Neural Networks - The Long Short-Term Memory and Other Gated RNNs

Unit:4	Deep Generative Models	12 hours
Deep Generative Models: Boltzmann Machines - Restricted Boltzmann Machines - Introduction to MCMC and Gibbs Sampling- gradient computations in RBMs - Deep Belief Networks- Deep Boltzmann Machines Applications: Large-Scale Deep Learning - Computer - Speech Recognition - Natural Language Processing - Other Applications		
Unit: 5	**Deep Learning: Practice**	**12 hours**
Deep Learning for Computer Vision – Training convents – Pretrained convnet – Visualizing convnet – Working with text data – Using word embeddings – Functional API – Text generation with LSTM – Implementing Deep Dream in Keras.		
Unit:6	**Contemporary Issues**	**2 hours**
Expert lectures, online seminars – webinars		
	Total Lecture hours	**60 hours**

Text Books

1	Abraham Silberchatz, Henry K.Forth, Sudharshan, "Database system Concepts", 7th edition, McGraw Hill, 2020.
2	Prabu C.S.R, "Object-Oriented Database Systems: Approaches and Architectures" 3rd Edition, PHI, 2011.
3	Kristina Chodorow , "MongoDB: The Definitive Guide", 3rd Edition , O'Reilly Media, ISBN: 9781491954461, 2019.
4	Guy Harrison, "Next Generation Databases: NoSQL, NewSQL, and Big Data",Apress, 2016.

Reference Books

1	ShamkantB.Navathe, RamezElamsri"Fundamentals of Database Systems", 7th Edition, Pearson Education Limited, 2017.
2	David Hows , Peter Membrey , EelcoPlugge , Timm Hawkins , "The Definitive Guide to MongoDB", 3rd Edition, Apress, 2015.
3	GauravVaish ,"Getting Started with NoSQL"Packt Publishing, 2013.
4	Ian Goodfellow, YoshuaBengio, Aaron Courville, "Deep Learning", MIT Press, 2016.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1. Deep Learning Specialization, https://www.coursera.org/specializations/deep-learning

Course Designed By: Prof. Dr. V. Bhuvaneswari

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	-	M	M	-	-	-	-	M	-	-
CO2	S	S	S	S	M	S	-	M	M	M
CO3	M	M	M	S	-	-	-	M	-	M
CO4	S	S	S	S	M	M	S	L	M	S
CO5	S	S	S	S	S	S	S	S	S	S

*S-Strong; M-Medium; L-Low

		SOCIAL MEDIA MINING	L	T	P	C
Core/Elective/Supportive		**Elective**	**4**			**4**

Pre-requisite	Nil	Syllabus Version	2023-2024

Course Objectives:

The main objectives of this course are to:

1. To understand how accurately analyze voluminous complex data set in social media and other sources
2. To understand the models and algorithms to process large data sets
3. To understand social behavior and recommendation challenges and methodologies

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

1	Understand the concepts of Graph Models, social communities	K1, K2
2	Understand the network models and measures to evaluate information	K3
3	Understand and apply algorithms to model data using graph and network structures and recommendations	K2,K5
4	Brief on algorithms on social data diffusion and apply for various domains	K2,K3, K4
5	Distinguish and Suggest the appropriate algorithms for domain specific applications for data modelling and information diffusion, Evaluate the algorithms for metrics	K4,K5, K6

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	Social Media Mining	12-- hours

Social Media Mining - Introduction – Atoms – Molecules – Interactions – Social Media mining Challenges - Graphs - Basics – Nodes – Edges – Degree of Distribution- Types –Directed – Undirected – Weighted - Graph Connectivity - Tress and Forests – Bipartite graphs – Complete Graphs – Sub graphs – Planar Graphs - Graph Representation - Graph Traversal Algorithms – Shortest path algorithms Dijkstra's - Spanning tree algorithms – Prims - Bipartite matching - Ford-Fulkerson algorithm

Unit:2	Network Models	12-- hours

Network Models – Measures – Node : Eigen Centrality – Page Rank – Group Measures – Betweenness centrality - group degree centrality, centrality, and group - Closeness centrality - Node Linking Behavior - Transitivity and reciprocity - Linking Analysis - Cluster coefficient – Jaccard - Case Study : -Modeling small networks with real world model

Unit:3	Social Media Communities	12-- hours

Social media Communities – Social Communities – Member based Detection – Node degree – Node Similarity – Node reachability - Group Based detection methods - balanced – robust - modular – dense - hierarchical - Spectral Clustering : Balanced Community algorithm Community Evolution - Evaluation.

Unit:4	Social Network	10-- hours

Social Network – Information Diffusion – Types - herd behavior - information cascades diffusion of innovation – epidemics – Diffusion Models Case Study – Herd Behavior – Information Cascades Methods – Social Similarity – assortativity – Social Forces - Influence homophily – Confounding - Assortativity measures – Influence measures – Predictive Models

Unit:5	Recommender System	12-- hours

Recommendation Vs Search – Recommendation Challenges – Recommender algorithms - Content-Based Methods- Collaborative Filtering – Memory Based – Model Based – Social Media Recommendation – User friendship – Recommendation Evaluation – Precision – Recall – Behavioral– User Behavior – User – Community behavior – User Entity behavior – Behavioral Analytics - Methodology

Unit:6	Contemporary Issues	2 hours

Expert lectures, online seminars – webinars

	Total Lecture hours	**60 hours**

Text Book(s)	
1	Reza Zafarani , Mohhammad AliAbbasi – Social Media Mining: An Introduction – Published by Cambridge press, 2014 – (Free Ebook available http://dmml.asu.edu/smm/chapter**)**
2	**Memon**, N., **Xu**, J.J., **Hicks**, D.L., **Chen**, H. (Eds.), Data Mining for Social Network Data- Springer – Annals of Information Systems ,ISBN 978-1-4419-6287-4
3	Lam Thuy Vo, 2019, "Mining Social Media: Finding Stories in Internet Data
Reference Books : EBooks	
1	Matthew A. Russel and Mikhail Klassen, 2018, "Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub
2	GungorPolatkan, AntonoisChalkiopoulos, P. Oscar Boykin et.al., 2018, "Social Media Mining and Analytics.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]			
	Course Title	**Duration**	**Provider**
1.	Social Media Data Analytics (Free)	4 Weeks	Coursera
2.	Introduction to Social Media Analytics	4 Weeks	Coursera
3.	Social Media Analytics: Using Data to Understand Public Conversations	3 Weeks	Future Learn
4.	Starting with social network analysis	2 hours	Udemy
Web link			

1. https://learn.g2.com/social-media-data-mining
2. https://www.javatpoint.com/social-media-data-mining
3. https://www.igi-global.com/dictionary/applying-critical-theories-to-social-media-mining-and-analysis/50376
4. https://www.cambridge.org/core/books/social-media-mining/introduction/75F143896832B7B9339F2CE663C4815B

Mapping with Programme Outcomes

Cos	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**
CO1	M	M	L	-	-	-	S	S	L	-
CO3	L	S	-	M	S	-	M	L	M	L
CO3	S	M	M	-	M	L	L	-	-	M
CO4	-	L	M	S	L	-	-	-	-	-
CO5	S	M		-	-	L	-	L	M	S

*S-Strong; M-Medium; L-Low

		VIRTUALIZAITON AND CLOUD	**L**	**T**	**P**	**C**
Core/Elective/Supportive		Elective	**4**			**4**
Pre-requisite		Basic knowledge of data storage, Client – Server systems	**Syllabus Version**		**2023-2024**	
Course Objectives:						

The main objectives of this course are: SCAA DATED: 18.05.2023

1. To impart knowledge on the concepts of distributed systems, cloud computing and AWS
2. To gain knowledge over various virtualization and virtual machines
3. To gain understanding about the data centers

Expected Course Outcomes:

On the successful completion of the course, student will be able to:

CO1	Understand the fundamentals of distributed systems	K2
CO2	Understand and use the cloud services and AWS	K3
CO3	Understand and perform virtualization	K3, K6
CO4	To create, configure and manage virtual machines	K4
CO5	Understand about data center	K5

K1 - Remember; **K2** - Understand; **K3** - Apply; **K4** - Analyze; **K5** - Evaluate; **K6** – Create

Unit:1	**Distributed Systems**	**12 hours**

Introduction to distributed systems - Distributed algorithm - Distributed Data Stores - Distributed Computing - File Systems - Distributed Messaging - Distributed Applications – Distributed Transaction - Parallel and distributed computing - Applications.

Unit:2	**Cloud Computing**	**12 hours**

Cloud Concepts: Introduction Cloud Computing - Advantages of Cloud - Public Cloud - five essential characteristics - three service models - Four deployment models - Benefits of Cloud Computing - Cloud Vendors - Traditional Infrastructure setup and Challenges – AWS.

Unit:3	**Virtualization**	**12 hours**

Virtualization: Introduction to vSphere and the Software - Defined Data Center - Creating Virtual Machines - VCenter Server - Configuring and Managing - Virtual Networks - Configuring and Managing Virtual Storage - Virtual Machine Management - Resource Management and Monitoring.

Unit:4	**Virtual Machines**	**12 hours**
Virtual Machines: vSphere HA - vSphere Fault Tolerance - Protecting Data vSphere DRS - Network Scalability - vSphere Update Manager and Host Maintenance - Storage Scalability - Securing Virtual Machines.		
Unit:5	**Datacentre**	**10 hours**
Data centre: Data centre overview -Components - Provisions - Need of Data Centre - Data Centre Architecture - Different Racks - Data center architecture for cloud computing - role of data centre in cloud computing.		
Unit:6	**Contemporary Issues**	**2 hours**
Expert lectures, online seminars – webinars		
	Total Lecture hours	**60 hours**

Text Book(s)

1	George Coulouris, Jean Dollimore, Tim Kindberg, Gordan Blair, "Distributed Systems Concepts and Design", 5thEdition, Pearson Education, 2012.
2	VenkataJosyula , Malcolm Orr , Greg Page, "Cloud Computing: Automating the Virtualized Data Center", 1st Edition, Cisco Press, 2011.
3	Brian J.S. Chee, Curtis Franklin Jr., "Cloud Computing: Technologies and Strategies of the Ubiquitous Data Center", 1st Edition, CRC Press, 2010.

Reference Books

1	Andrew S. Tanenbaum, Maarten Van Steen, "Distributed Systems: Principles and Paradigms", 2nd edition, Createspace Independent Publishers, 2016.
2	Matthew Portnoy, "Virtualization Essentials", 2nd edition, Wiley Publication, 2016.

Related Online Contents [MOOC, SWAYAM, NPTEL, Websites etc.]

1	Cloud Computing and Distributed Systems, https://nptel.ac.in/courses/106/104/106104182/

Mapping with Programme Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	S	S	-	-	-	-	M	-	-	-
CO2	-	S	M	S	-	-	-	-	-	-
CO3	-	M	S	L	-	-	L	S	-	M
CO4	-	L	S	M	-	-		M	-	M
CO5	-	L	S	S	-	-	M	M	-	L

*S-Strong; M-Medium; L-Low