ST. XAVIER'S COLLEGE (AUTONOMOUS)
Palayamkottai - 627 002
Recognized as "College with Potential for Excellence" by UGC
Accredited at A++ Grade with a CGPA of 3.66 out of 4 in IV Cycle by NAAC



SYLLABUS

M.Sc. DATA SCIENCE
(W.e.f. June 2023)

Programme Name: M.Sc. Data Science

Programme Code: PDS

Programme Outcomes (POs) for PG Programmes

Students of all Postgraduate Degree Programmes at the time of graduation will be able to attain the following at the institution level:

PO1: Critical Thinking: Acquire the knowledge in the respective field and take informed actions.

PO2: Communication: Speak, read, write, and listen clearly in person and through electronic media in English and in one Indian language, and make meaning of the world by connecting people, ideas, books, media, and technology.

PO3: Individual and teamwork: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO4: Problem analysis: Identify, formulate, review research literature, and analyze complex problems reaching substantiated conclusions using the principles of all branches of sciences, commerce, economics, management studies, language, and literature.

PO5: Design/development of solutions: Design solutions for problems and design system components or processes that meet the specified needs with appropriate consideration for the public, health and safety, and the cultural, societal, and environmental considerations.

PO6: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO7: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern IT tools including prediction and modeling to complex activities with an understanding of the limitations.

PO8: Self-directed and Life-long Learning: Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of socio-technological changes.

Programme Name: M.Sc. Data Science
Programme Code: PDS

Program Specific Outcomes (PSOs):

Students of Postgraduate Degree Programme of Data Science at the time of graduation will be able to attain the following at the department level:

PSO1: Domain Knowledge: Apply the knowledge of all courses for the solution of data science problems.

PSO2: Apply Concepts in Real World Problems: Ability to analyze and implement the concepts and techniques in the field of data science to solve real world problems.

PSO3: Impart Entrepreneurial Skills: Analyze Complex Problems and system components through visualization of data in different forms that will pave the way for inventing different products in the field of data science.

PSO4: Analysis and Evaluation of Procedures: Use research-based knowledge and research methods including data exploration, statistical analysis, design of experiments, machine learning and performance analysis to provide valid conclusions.

PSO5: Contribution to Business World and the Society: Create innovative solutions ethically to the problems in industry and society to envisage a better world.

PG Data Science Programme Pattern
(With Effect from June 2023)
DATA SCIENCE

Sem	Part	Status	Sub. Code	Title of the Paper	Hrs	Cdt
I	A	Core-1	23PDSC11	Fundamentals of Data Science	7	5
		Core-2	23PDSC12	Mathematics for Data Science	7	5
		Core-3	23PDSC13	Statistics – I	6	4
		EC-1	23PDSE11	Data Structures & Algorithms/Internet of Things/ Research Methodology for Data Science	5	3
		EC-2	23PDSE12	Java Programming/ Web Programming/ Operating Systems (Linux Based Commands for Practical's)	3	2
		EC-3	23PDSE13	Java Programming Lab/ Web Programming Lab/ Operating Systems Lab	2	1
					30	**20**
II	A	Core-4	23PDSC21	Python Programming	6	5
		Core-5	23PDSC22	Python Programming Lab	6	5
		Core-6	23PDSC23	Statistics – II	6	4
		EC-4	23PDSE21	Software Engineering for Data Science/Distributed Systems/ Information Security and Ethics	4	3
		EC-5	23PDSE22	Discrete Mathematics/Applied Probability/ Optimization Techniques	4	3
	B	SEC2	23PDSS21	Emerging Technologies in Data Science	4	2
					30	**22**
III	A	Core-7	23PDSC31	Machine Learning	6	5
		Core-8	23PDSC32	Machine Learning Lab	6	5
		Core-9	23PDSC33	Data Bases for Data Science	6	5
		Core-10	23PDSC34	Natural Language Processing	6	4
		EC-6	23PDSE31	Predictive Data Analytics/Social Network Analytics/ Reinforcement Learning/ Business and Healthcare Analytics	4	3
	B	SEC-2	23PDSS31	Professional Communication Skill	2	2
		Internship	23PDSI35	Carried out in summer vacation at the end of Sem II	-	2
					30	**26**
IV	A	Core-11	23PDSC41	Deep Learning	6	5
		Core-12	23PDSC42	Big Data Analytics and Big Data Analytics Lab	6	5
		Project	23PDSC43	Project With Viva Voce	10	7
		EC-7	23PDSE41	Artificial Intelligence/ Image Analytics/ Cloud Computing/ Recommendation System	4	3
	B	SEC-3	23PDSS41	Deep Learning Lab	4	2
	C	Extension Activity		STAND carried out in the I year	-	1
					30	**23**
				Total	**120**	**91**

ADD ON COURSES

Sem	Status	Sub. Code	Title of the Paper	Hrs	Cdt
I	ADC –1	23PDSAO1	Excel(Basic Level)	35	3
	ADC –2	23PDSAO2	Excel(Intermediate Level)	35	3
II	ADC –3	23PDSAO3	R (Basic Level)	35	3
	ADC –4	23PDSAO4	R (Intermediate Level)	35	3
III	ADC –5	23PDSAO5	Python for Data Science	35	3
	ADC –6	23PDSAO6	Natural Language Processing Lab	35	3
IV	ADC –7	23PDSAO7	Power BI	35	3
	ADC –8	23PDSAO8	Tableau	35	3

Value Added Courses

Sem	Status	Sub. Code	Title of the Paper	Hrs	Cdt
I	VAC – 1	23PDSVA1	Excel for Statistical and Data Analytics	40	3
II	VAC – 2	23PDSVA2	Database Lab for Data Science (Basic Level)	40	3
III	VAC – 3	23PDSVA3	Database Lab for Data Science (Advanced Level)	40	3
IV	VAC – 4	23PDSVA4	Cloud Computing Lab	40	3

Certificate Course

Sem	Status	Sub. Code	Title of the Paper	Hrs	Cdt
I	CEC – 1	23PDSCC1	Java Programming Lab (Advanced Level)	30	3
	CEC – 2	23PDSCC2	Web Programming	30	3
II	CEC – 3	23PDSCC3	MySQL	30	3
	CEC – 4	23PDSCC4	NoSQL	30	3
III	CEC – 5	23PDSCC5	Business Analytics Lab	30	3
	CEC – 6	23PDSCC6	Health Care Analytics	30	3
IV	CEC – 7	23PDSCC7	Multimedia I (Other Department Students)	30	3
	CEC – 8	23PDSCC8	Multimedia II (Other Department Students)	30	3

Extra Credit Course

Sem	Status	Sub. Code	Title of the Paper	Hrs	Cdt
I	ECC – 1	23PDSEC1	Microsoft Excel(Other Department Students)	30	3
II	ECC – 2	23PDSEC2	Microsoft Office Tools (Other Department Students)	30	3
III	ECC – 3	23PDSEC3	Web Scraping using Python (Data Science Students)	30	3
IV	ECC – 4	23PDSEC4	Financial Analytics using Excel (Data Science Students)	30	3

Abbreviations

EC – Elective Course
SEC – Skill Enhancement Course

FUNDAMENTALS OF DATA SCIENCE
(Course code: 23PDSC11)

SEMESTER– I	HOURS –7	CREDITS – 5

Objective:

To be able to rehearse the fundamental data science concepts, apply mathematical principles to experiment diverse data-intensive tools, and apply appropriate methods to create innovate solutions for solving real-world problems.

Outcomes:

At the end of the course the students will be able to:

CO1: Describe the fundamental concepts and principles of data science and how they are related to real-world scenarios. (K3)

CO2: Associate the mathematical knowledge with data science by solving related problems (K2)

CO3: Experiment various data intensive operations and tools. (K3)

CO4: Examine the effectiveness and relevance of various tools and techniques for analyzing and visualizing data. (K4)

CO5: Criticize the recent potential applications while selecting techniques and tools to analyze real time problems(K5)

CO6: Formulate pioneering data-driven solutions for real-world challenges by selecting tools and techniques, and visualization strategies while relating them with the recent trends. (K6)

UNITI: (14 hrs)

INTRODUCTION OF DATA SCIENCE: Data Science –Venn diagram – Basic terminology – Data science case studies – Types of data – levels of data – Types of data analytics – Descriptive analytics–Diagnostic analytics– Predictive analytics- Prescriptive analytics- Five steps of Data science.

UNIT II: (15 hrs)

MATHEMATICAL PRELIMINARIES: Basic Maths – mathematics as discipline – basic symbols and terminology –linear algebra. Basic Probability – definitions- probability – Bayesian vs frequentist – compound events – conditional probability – rules of probability.

UNIT III: (14hrs)

DATA MINING AND DATA WAREHOUSING: Introduction to Data warehousing – Design consideration of data warehouse – Data loading process – case study – Data mining – Data mining techniques – Tools and platforms – case study.

UNIT IV: (15 hrs)

VISUALIZING DATA: Exploratory Data Analysis – Developing the visual aesthetic – chart types – Great visualizations – Reading graphs – Interactive visualizations.

UNIT V: (14 hrs)

Data Science – Recent Trends: Applications of Data Science, recent trends in various data collection and analysis techniques, various visualization techniques, application development methods of used in data science.

Text Books:

1. Ozdemir, Sinan, Principles of data science. Packt Publishing Ltd, 2016. (**Unit 1**- Ch. 1,2,3 **Unit II** – Ch. 4, Ch. 5)

2. Maheshwari, Anil, "Data Analytics Made Accessible." Seattle: Amazon Digital Services, 2nd edition, 2023.(**Unit III** – Ch. 3, Ch. 4)
3. Skiena, Steven S. The data science design manual. Springer, 2017. (**Unit IV**- Ch. 6)

Reference Books:

1. Hadrien Jean C., Data Science,Certybox Education, 2023.
2. Pierson, Lillian, Data Science for Dummies, John Wiley & Sons, 2021.
3. Grus, Joel, Data Science from Scratch: First Principles with Python, O'Reilly Media, 2019.
4. Blum, Avrim, John Hopcroft, Ravindran Kannan, Foundations of Data Science, Cambridge University Press, 2020.

Web References:

- https://www.analyticsvidhya.com/
- https://www.simplilearn.com
- https://www.ibm.com/in-en/topics/data-science
- https://www.mygreatlearning.com/blog/what-is-data-science/

MATHEMATICS FOR DATA SCIENCE
(Course code: 23PDSC12)

SEMESTER– I	HOURS – 7	CREDITS – 5

Objective:

To build the mathematical background necessary to understand and implement in data science practical/research work.

Outcomes:

At the end of the course the students will be able to:

CO1: Demonstrate understanding of basic mathematical concepts in Linear Algebra which are needed for data science. (K3)

CO2: Analyze the properties of linear systems using matrices/vectors to perform and interpret matrix operations. (K4)

CO3: Illustrate the relationships and interactions between the Four Fundamental subspaces. (K4)

CO4: Outline the properties in orthogonality, determinants, Eigen values, Eigen vectors in Linear Algebra. (K4)

CO5: Justify the relationship and interaction between the different types of procedures and theories learned in Eigen values, Eigenvectors and Linear Differential Equations. (K5)

CO6: Summarize the properties and the principles learned in Singular Value Decomposition and Linear Transformation. (K5)

UNIT-I: (16 hrs)

Vectors and Matrices: Vectors and Linear Combinations-Lengths and Angles from Dot Products-Matrices and Their Column Spaces-Matrix Multiplication AB and CR. **Solving Linear Equations Ax = b:** Elimination and Back Substitution-Elimination Matrices and Inverse Matrices-Matrix Computations and A = LU-Permutations and Transposes.

UNIT-II: (17 hrs)

The Four Fundamental Subspaces: Vector Spaces and Subspaces-Computing the Null space by Elimination: A = CR-The Complete Solution to Ax = b-Independence, Basis, and Dimension-Dimensions of the Four Subspaces.

UNIT-III: (16 hrs)

Orthogonality: Orthogonality of Vectors and Subspaces-Projections onto Lines and Subspaces-Least Squares Approximations - Orthonormal Bases and Gram-Schmidt-The Pseudo inverse of a Matrix. **Determinants:** 3 by 3 Determinants and Cofactors-Computing and Using Determinants-Areas and Volumes by Determinants.

UNIT-IV: (17 hrs)

Eigen values and Eigenvectors: Introduction to Eigenvalues: $Ax = \lambda x$– Diagonalizing a Matrix-Symmetric Positive Definite Matrices-Complex Numbers and Vectors and Matrices-Solving Linear Differential Equations.

UNIT-V: **(16 hrs)**

The Singular Value Decomposition (SVD): Singular Values and Singular Vectors-Image Processing by Linear Algebra-Principal Component Analysis (PCA by the SVD).

Linear Transformations: The Idea of a Linear Transformation–The Matrix of a Linear Transformation–The Search for a Good Basis.

Text Books:

1. Gilbert Strang, Introduction to Linear Algebra, Wellesley – Cambridge Press, 6th Edition, 2023.

Reference Books:

1. David Lay, Steven Lay, Judi McDonald, Linear Algebra and Its Applications, 5th Edition, Pearsons.
2. Sheldon Axler, Linear Algebra Done Right (Undergraduate Texts in Mathematics), 3rd ed., Springer, 2015.
3. Jim Hefferon, Linear Algebra, Fourth edition, OpenIntro, 2021.
4. Jeff M Philips, Mathematical Foundations for Data Analysis, Springer Cham, 2021.

Web References:

- https://joshua.smcvt.edu/linearalgebra/

STATISTICS - I
(Course code: 23PDSC13**)**

SEMESTER– I	HOURS – 6	CREDITS – 4

Objective:

To develop knowledge and understand fundamental concepts in probability and statistics.

Outcomes:

At the end of the course the students will be able to:

CO 1: Recite a comprehensive understanding of fundamental statistical concepts and their applications.(K2)

CO 2: Contra stand interpret data using appropriate descriptive statistical measures, graphical representations, and frequency distributions. (K2)

CO 3: Apply probability models and inferential statistics to make accurate inferences and predictions from sample data. (K3)

CO 4: Select appropriate sampling techniques and assess the reliability of sampling distributions. (K4)

CO 5: Recommend population parameters using maximum likelihood estimators and construct confidence intervals for various distributions. (K5)

CO 6: Derivesolution for real-world problems integrating probability, correlation, and data analysis. (K6)

UNIT-I: (18 hrs)

Introduction to Statistics: Introduction-Data Collection and Descriptive Statistics-Inferential Statistics and Probability Models-Populations and Samples-A Brief History of Statistics **Organization and Presentation of Data:** Origin and development of Statistics, Scope, limitation, and misuse of statistics. Types of data: primary, secondary, quantitative, and qualitative data. **Types of Measurements:** nominal, ordinal, discrete and continuous data. Presentation of data by tables: construction of frequency distributions for discrete and continuous data, graphical representation of a frequency distribution by histogram and frequency polygon, cumulative frequency distributions.

UNIT-II: (17 hrs)

Descriptive statistics: Introduction-Describing Data Sets-Frequency Tables and Graphs-Relative Frequency Tables and Graphs-Grouped Data, Histograms, Ogives, and Stem and Leaf Plots-Summarizing Data Sets-Sample Mean, Sample Median, and Sample Mode-Sample Variance and Sample Standard Deviation-Sample Percentiles and Box Plots-Chebyshev's Inequality-Normal Data Sets-Paired Data Sets and the Sample Correlation Coefficient. **Correlation:** Scatter plot, Karl Pearson coefficient of correlation, Spearman's rank correlation coefficient, multiple and partial correlations (for 3 variates only).

UNIT-III: (17 hrs)

Random variables and expectation: Random Variables-Types of Random Variables-Jointly Distributed Random Variables-Independent Random Variables-Conditional Distributions-Expectation-Properties of the Expected Value-Expected Value of Sums of Random Variables-Variance-Covariance and Variance of Sums of Random Variables-Moment Generating Functions-Chebyshev's Inequality and the Weak Law of Large Numbers. **Special**

random variables: The Bernoulli and Binomial Random Variables-Computing the Binomial Distribution Function-The Poisson Random Variable-Computing the Poisson Distribution Function-The Hypergeometric Random Variable-The Uniform Random Variable- Normal Random Variables-Exponential Random Variables-The Poisson Process-The Gamma Distribution-Distributions Arising from the Normal-The Chi-Square Distribution-The t-Distribution-The F Distribution-The Logistics Distribution**.**

UNIT-IV: (17 hrs)

Distributions of sampling statistics: Introduction-The Sample Mean-The Central Limit Theorem-Approximate Distribution of the Sample Mean, How Large a Sample Is Needed? – The Sample Variance-Sampling Distributions from a Normal Population-Distribution of the Sample Mean, Joint Distribution of X and S-Sampling from a Finite Population. **Parameter estimation:** Introduction-Maximum Likelihood Estimators-Interval Estimates-Confidence Interval for a Normal Mean When the Variance is Unknown-Confidence Intervals for the Variance of a Normal Distribution – Estimating the Difference in Means of Two Normal Populations-Approximate Confidence Interval for the Mean of a Bernoulli Random Variable-Confidence Interval of the Mean of the Exponential Distribution-The Bayes Estimator.

UNIT-V: (17 hrs)

Basics and Elements of Probability: Random experiment, sample point and sample space, event, algebra of events. Definition of Probability: classical, empirical and axiomatic approaches to probability, properties of probability. Theorems on probability, conditional probability and independent events, Laws of total probability, Baye's theorem and its applications-Introduction-Sample Space and Events-Venn Diagrams and the Algebra of Events-Axioms of Probability-Sample Spaces Having Equally Likely Outcomes.

Text Books:

1. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 11th edition, Sultan Chand & Sons, New Delhi, 2014. (**Unit I)**

2.Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, Elsevier Academic Press, UK, Fifth Edition, 2023. (**Unit II, Unit III**, **Unit V**)

3.Rohatgi V.K and Saleh E, An Introduction to Probability and Statistics, 3rd edition, John Wiley & Sons Inc., New Jersey, 2015. (**Unit IV)**

Reference Books:

1. Jim Frost, Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries.

Web References:

- https://onlinestatbook.com/2/
- https://www.simplilearn.com/tutorials/statistics-tutorial
- https://towardsdatascience.com/fundamentals-of-statistics-for-data-scientists-and-data-analysts-69d93a05aae7

DATA STRUCTURES AND ALOGORITHMS
(Course code: 23PDSE11)

SEMESTER– I	HOURS –5	CREDITS – 3

Objective:

To develop an understanding of the research methods relevant to effectively address a research problem.

Outcomes:

At the end of the course the students will be able to:

CO1 : Recite the concepts related to analysing and designing algorithms using different problem - solving techniques. (K1)

CO2 : Differentiate and implement abstract data types (ADTs) and data structures to solve real-world problems. (K2)

CO3 : Experiment different techniques related to efficient data storage and retrieval. (K3)

CO4 : Analyse the efficiency of algorithms and techniques. (K4)

CO5 : Evaluate and select graph-related problems using graph traversal algorithms and minimum spanning trees. (K5)

CO6: Develop algorithms and compute the complexity while classifying them as NP-complete or NP-hard. (K6)

UNIT-I: **(12 hrs)**

Basic Concepts: Basic steps in complete development of Algorithm –Analysis and complexity of Algorithm – Asymptotic notations – Problem Solving techniques and examples. **ADT:**

List ADT, Stacks ADT, Queue ADT.

UNIT-II: **(12 hrs)**

Algorithm Design Model: Greedy Method – Divide and Conquer – Dynamic Programming – Backtracking – Branch and Bound. **Trees:** Preliminaries Binary Tree–Search Tree ADT–Binary Search Trees–AVL Trees–Tree Traversals–B-Trees.

UNIT-III: **(12 hrs)**

Hashing: General Idea – Hash Function – Separate Chaining – Open Addressing - Rehashing – Extendible Hashing – Priority Queues – Model – Simple Implementations – Binary Heap – Applications.

UNIT-IV: **(12 hrs)**

Sorting: Sorting – Preliminaries–Insertion Sort–Shell Sort–Heap Sort–Merge Sort–Quick Sort–External Sorting.

UNIT-V: **(12 hrs)**

Graphs: Definitions, Topological Sort, Shortest Path Algorithm, Minimum Spanning Tree, Application of Depth First Search. **Theory of NP-Completeness:** Formal language framework, Complexity classes – P, NP – NP Reducibility and NP-Complete, NP-Hard.

Text Books:

1. Aho, J. E. Hopcroft, J. D. Ullman, Design and Analysis of Computer Algorithms. 1st ed, Addison-Wesley, 2009.
2. Horowitz and Sahani, Fundamentals of Computer Algorithms, 2nded,Galgotia, 2008.
3. Weiss, M. A, Data Structure and Algorithm analysis in C, 2nd ed, Pearson Education Asia, 2002.

Reference Books:

1. Baase, S., Allen Van Gelder, Computer Algorithms-Introduction to Design and Analysis, New Delhi, Pearson Education, 2008
2. Goodrich, M.T., R. Tamassia. Algorithm Design: Foundations, Analysis, and Internet Examples. New Delhi: Wiley, 2006.

Web References:

- https://www.geeksforgeeks.org/learn-data-structures-and-algorithms-dsa-tutorial/
- https://shorturl.at/cjmoN

INTERNET OF THINGS
(Course code: 23PDSE11)

SEMESTER– I	HOURS –5	CREDITS –3

Objective:

To understand the concepts, data, framework, standards, protocols, reliability, security, and privacy involved in IoT.

Outcomes:

At the end of the course the students will be able to:

CO1: Identifyprotocols, architectures, deployment models and their applications. (K1)
CO2: Associate IoT solutions using programming frameworks and devices. (K2)
CO3: Examine security frameworks and implement measures for IoT security. (K3)
CO4: Categorize and select protocols and communication standards for IoT networks.(K4)
CO5: Assess IoT data management and analytics techniques for efficient processing. (K5)
CO6: Design and integrate frameworks for effective IoT management. (K6)

UNIT-I: (12 hrs)

IoT Ecosystem Concepts and Architectures: Introduction – IoT definition and evolution – IoT Architectures –OpenIoT Architecture for IoT/Cloud Convergence – Resource Management – IoT Data Management and Analytics – Communication Protocols – Applications – SchedulingProcess and IoT Services Lifecycle – IoT enabling technologies – IoT levels and Deployments templates – Introduction to M2M – Difference between IoT and M2M – SDN and NFV for IoT.

UNIT-II: (12 hrs)

IoT Data and Framework Essentials –Programming framework for IoT– The foundation of Stream processing in IoT – Continuous Logic processing system – Challenges and Future directions – Anomaly detection – Problem statement and definitions – Efficient incremental local modelling – IoT Governance.

UNIT-III: (12 hrs)

RF Protocols RFID, NFC;IEEE 802.15.4: Zig Bee – ZWAVE, THREAD – Bluetooth Low Energy (BLE) – Ipv6 for Low Power and Lossy Networks (6LoWPAN) – Routing Protocol for Low power and lossy networks (RPL) – CoAP – XMPP – Web Socket- AMQP – MQTT – WebRTC –PuSH Architectural Considerations in Smart Object Networking –TinyTO Protocol. IoT based applications – Scenarios – Architecture overview – Sensors – The gateway – Data Transmission – Internet of Vehicles (IoV) – IoV Characteristics, technologies and its application.

UNIT-IV: (12 hrs)

Developing Internet of Things: IoT Design Methodology – Case study on IoT system for Weather monitoring – IoT Device – IoT physical devices and endpoints – Exemplary Device: Raspberry Pi – Linux on Raspberry Pi – Raspberry Pi interfaces – Programming Raspberry Pi and with python – Other IoT devices.

UNIT-V: (12 hrs)

IoT Reliability, Security and Privacy: Concepts – IoT Security Overview – Security Frameworks for IoT – Privacy in IoT networks – IoT characteristics and reliability issues – Addressing reliability.

Text Books:

1. Arshdeep Bahga, Vijay Madisetti, "Internet of Things, A Hands -on Approach", 1st Edition 2015, University Press, ISBN: 978-81-7371- 954-7
2. Buyya, Rajkumar, Amir Vahid Dastjerdi, eds. Internet of Things: Principles and paradigms. Elsevier, 2016.
3. Hersent, Olivier, David Boswarthick, Omar Elloumi. The internet of things: Key applications and protocols. John Wiley & Sons, 2011.

Reference Books:

1. Bernd Scholz-Reiter, Florian Michahelles, "Architecting the Internet of Things", ISBN 978- 3-642-19156-5 e-ISBN 978-3-642-19157-2, Springer
2. Jan Holler, Vlasios Tsiatsis, Catherine Mulligan, Stefan Avesand, Stamatis Karnouskos, David Boyle, "From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence", 1st Edition, Academic Press, 2014.
3. Peter Waher, "Learning Internet of Things", PACKT publishing, Birmingham – Mumbai.

Web References:

- https://thingsee.com/blog/quality-hardware-list-for-your-iot-projects
- https://tools.ietf.org/html/rfc7452.
- http://dret.net/lectures/iot-spring15/protocols
- http://iot.intersog.com/blog/overview-of-iot-development-standards-andframeworks.

RESEARCH METHODOLOGY FOR DATA SCIENCE
(Course code: 23PDSE11)

SEMESTER– I	HOURS –5	CREDITS –3

Objective:

To develop an understanding of the research methods relevant to effectively address a research problem.

Outcomes:

At the end of the course the students will be able to:

CO1: Explain the fundamental concepts and characteristics of research, including its objectives and the criteria for good research. (K2)

CO2: Differentiate between research methods and methodology and apply appropriate research methods to different types of research. (K2)

CO3: Develop research projects by effectively shaping and planning them, considering the roles of students and advisors. (K3)

CO4: Conduct a comprehensive literature review, including reading, reviewing, and synthesizing relevant information, and formulating hypotheses, questions, and evidence. (K4)

CO5: Demonstrate proficiency in writing research papers, including organizing content, adhering to good style guidelines, using proper punctuation, incorporating mathematical equations and algorithms, and presenting graphs, figures, and tables effectively. (K5)

CO6: Prepare professional presentations, including editing content, creating slides and posters, and adhering to ethical guidelines, and utilize LaTeX for report writing on a research problem. (K6)

UNIT-I: (15 Hours)

Introduction to Research: Meaning, Objectives and Characteristics of research - Research Methods Vs. Methodology - Types of research- Research process - Criteria of good research.

UNIT-II: (15 Hours)

Research Project: Shaping a Research Project-Research Planning-Students and Advisors – Checklist.
Literature Review: Reading and Reviewing - Hypotheses, Questions, and Evidence.

UNIT-III: (15 Hours)

Experiments for Computing: Experimentation-Statistical Principles. **Writing a Paper:** Organization-Good Style-Style Specifics-Punctuation-Mathematics-Algorithms- Graphs, Figures, and Tables -Other Professional Writing.

UNIT-IV: (15 Hours)

Presentation: Editing – Presentations – Slides – Posters – Ethics.

UNIT-V: (15 Hours)

Report writing: Report writing using LATEX for a research problem.

Text Books:

1. Kothari C. R. Research Methodology Methods and Techniques. 2nd ed. New Delhi: New Age, 2004.
2. Justin Zobel. Writing for Computer Science.3rd ed. Springer-Verlag,2014

Reference Books:

1. Ranjit Kumar. Research Methodology -a step-by-step guide for beginners. 3rd ed. SAGE Publications India Pvt Ltd, 2011.
2. Panneerselvam R. Research Methodology. 2nd ed. New Delhi: Prentice Hall, 2014.

Web References:

1. https://www2.le.ac.uk/offices/red/rd/research-methods-and-methodologies
2. http://www.socscidiss.bham.ac.uk/methodologies.html

JAVA PROGRAMMING
(Course code : 23PDSE12)

SEMESTER– I	HOURS –5	CREDITS –3

Objective:

To enable the students to understand and appreciate the need for Object Oriented Programming and the use of java programming constructs.

Outcomes:

At the end of the course the students will be able to:

CO1: Determine the concepts of object-orientation and java programming. (K3)
CO2: Associate object-oriented concepts in Java for developing applications. (K2)
CO3: Illustrate the use of exceptions and perform database connectivity in Java. (K3)
CO4: Analyze the functionality of inheritance, interface for developing java applications. (K4)
CO5: Evaluate java packages, classes and collections and choose them for building apps. (K5)
CO6: Develop apps based on interactive user interfaces utilizing AWT controls.(K6)

UNIT-I: (15 hrs)

Introduction to Java: Overview – Features – Fundamental OOPS concepts – JDK – JRE – JVM -Structure of a Java program – Data types – Variables – Arrays – Operators –Keywords – Naming Conventions – Control statements, Type conversion and Casting – Scanner – String –equals(), equals Ignore Case(), length().

UNIT-II: (15 hrs)

Classes and Objects: Class – Objects – Methods – Method Overloading – Constructors– Constructor Overloading – this keyword – static– Garbage Collection – Access Control. **Inheritance:** Concept – extends keyword – Single and Multilevel Inheritance – Composition – super keyword – Method Overriding – Abstract Classes –Dynamic Method Dispatch –final. **Packages and Interfaces:** Concepts – package and import keywords – Defining, Creating and Accessing a Package – Interfaces – Multiple Inheritance in Java - Extending and Initialising fields in Interfaces.

UNIT-III: (15 hrs)

Exception Handling: Exception handling- Types of Exceptions- try, catch, throw, throws and finally keywords – User defined Exceptions. **JDBC:** Database Connectivity- Types of JDBC drivers- Executing statements- Prepared statements- Callable statements – Mapping SQL types to Java- Result Set Metadata.

UNIT-IV: (15 hrs)

Multithreading: Introduction – Life Cycle of a Thread, Thread class and Runnable Interface, Thread Priorities, Synchronisation. **GUI Programming with Java FX:** Java FX Basic Concepts – Packages – Stage and Scene Classes – Nodes and Scene Graphs – Layouts – The Application Class and the Lifecycle Methods – Launching a Java FX Application – Java FX Application Skeleton – Compiling and Running -Application Thread. **Java FX Controls:** Label – Button – Image – Radio Button – Check Box – List View- Combo Box- Text Field – Scroll Pane.

UNIT-V: (15 hrs)

Event: Event Handling – Input Event, Action Event and Window Event. **Java Library:** Java.util – List, Array List- Applets - **Advanced Java**: Remote Method Invocation - Swing: Frames – JApplets – JFrames – Servlets.

Text Books:

1. Schildt, Herbert. Java: The Complete Reference. McGraw-Hill Education Group, 2014.

Reference Books:

1.Eckel, Bruce. Thinking in Java. 4th ed. Pearson Education, 2006.
2. Liang, Y. Daniel. Intro to Java Programming, Brief Version. Pearson Higher Ed, 2015.
3. Holmes, J. Barry, Joyce, T. Daniel. Object-oriented Programming with Java. Jones & Bartlett Learning, 2001.

Web References:

- http://docs.oracle.com/javase/tutorial/java/index.html/
- http://www.java2s.com/Tutorial/Java/CatalogJava.htm/
- https://www.edureka.co/blog/object-oriented-programming/

WEB PROGRAMMING

(Course Code: 23PDSE12)

SEMESTER– I	HOURS –5	CREDITS –3

Objective:
To introduce students about web application and state management

Outcomes:
At the end of the course the students will be able to:
CO1: Contrast the components of .NET Framework and Windows Applications. (K2)
CO2: Restate the use of controls and events for developing Windows applications. (K2)
CO3: Examine the use of data types, namespaces, and ADO.NET for data manipulation. (K3)
CO4: Illustrate web server controls for building web applications. (K4)
CO5: Summarize state management techniques in ASP.NET and React Native. (K5)
CO6 : Create applications utilizing advanced features of .NET, and React Native. (K6)

UNIT-I: (12 hrs)
Overview of .NET Framework: CLR-CTS- Metadata and Assemblies-.NET Framework Class Library – BCL- Windows Forms – Tools in the .NET Framework- New Features of .NET Framework: Portable Class Libraries. **Introducing Windows Application:** Introduction – Creating Windows Forms- Customizing a Form - Collecting User Input in windows Forms and Events. Presentation and Informational Controls - **Data Types in C#:** Conversions – Boxing and Unboxing.

UNIT-II: (12 hrs)
Namespaces: Adding a reference to the Namespace – Accessing a predefined Namespace through the using Directive. **ADO.net:** Creating Connection Strings –Connection to a Database- Command Object- Data Adapters –Using DataReader.**ASP.NET**: Life cycle – Specifying Location– Single – File Page Model – Code-Behind Page Model- Adding controls to web form.**Web Server Controls**– Master Page –Web.Config.

UNIT-III: (12 hrs)
State Management: Sessions - Configuring - Session state - Application Object - Query strings – Cookies – ViewState - Global.asax.**LINQ:** Introducing LINQ Queries- Standard Query Operators- Introducing LINQ to Dataset, SQL and XML- Linq DataSource Control. Data Binding – Grid View - Details view - Forms view - **ASP. NET AJAX:** Need for AJAX - ASP.NET page with AJAX.

UNIT-IV: (12 hrs)
React Native: Advantages - Lifecycle- Using JSX- Styling Native Components- Host Platform APIs - Building Your First Application - Components for Mobile - Gesture Responder System- Pan Responder.

UNIT-V: (12 hrs)
Organizational Components - Platform-Specific Components - Styles – Layouts - Platform APIs – Modules - Deploying to the iOS App Store - Uploading - Beta Testing - Deploying Android Applications - Building the APK for Release - Submitting to the Play Store.

Text Books:

1. C# 2012 Programming Covers .NET 4.5 Black Book. Dreamtech press, Kogent Learning Solutions, 2013.(Unit 1)
2. Liberty, Jesse, and Dan Hurwitz. Programming. NET Windows Applications. "O'Reilly Media, Inc.", 2004. (Unit 2)

3. Troelsen, Andrew, and Philip Japikse, C# 6.0 and the .NET 4.6 Framework. Apress, 2015. (Unit 3)
4. Bonnie Eisenman, Learning React Native- Building Mobile Applications withJavaScript, O'Reilly Media, USA, 2016. (Unit IV, V)

Reference Books:

1. Albahan Joseph, Ben Albahari. C# 5.0 in a NutShell: The Definitive Reference. "Orielly Media Inc", 2012.
2. Anne Boehm, Joel, Murach's C# 2015. United States of America: Murach's,2016.
3. Delamater. Mary. Anne Boehm. ASP.NET 4.5 Web Programming with C# 2012. United Statesof America: Murach's, 2013.
4. Nader Dabit, "React Native in Action"- Developing iOS and Android apps withJavaScript, Manning Publications Co. USA, 2019

Web References:

- http://www.w3schools.com/aspnet/aspnet.asp
- http://csharp.net-tutorials.com/xml/introduction/
- http://ajax.net-tutorials.com/basics/introduction/
- http://www.c-sharpcorner.com/

OPERATING SYSTEMS
(Course code : 23PDSE12)

SEMESTER– I	HOURS –4	CREDITS – 3

Objective:
To learn the principles, architectures, Processes, Communication, Co-ordination, consistency, and Replication in Distributed Systems.

Outcomes:
At the end of the course the students will be able to:

CO1: Describe the fundamental concepts and components of operating systems, including their structure, operations, protection, security, and kernel data structures. (K1)

CO2: Demonstrate proficiency in using UNIX operating systems, including knowledge of the file system structure and basic UNIX commands for file and directory management. (K2)

CO3: Explain process management in operating systems, including concepts such as process control blocks, scheduling algorithms, CPU scheduling, process synchronization, and deadlock prevention and recovery. (K3)

CO4: Apply different file organization techniques, access methods, and file allocation methods, while considering file protection, security, directory structures, and efficiency in terms of performance and recovery. (K4)

CO5: Evaluate the memory management techniques, including single and multiple partition allocation, swapping, paging, segmentation, virtual memory, and address mapping, as well as device management and disk utilities. (K5)

CO6: Test the concepts related to operating system security, including controlling access to resources, cryptography, authentication, and defenses, as well as virtualization, cloud computing, and their associated security considerations. (K6)

UNIT-I: (8 hrs)

Introduction to Operating System: Structure – Operations - Protection and Security - Kernel Data Structures - Computing Environments – Services - System Calls and its types - System Programs - OS Design and Implementation - System Generation - System Boot. **UNIX Operating System:** History of UNIX, Shell, UNIX File System Structure.
Basic UNIX Commands: Commands for files and directories - creating and viewing files-using cat, date, who, pwd - filter commands –head tail, cut, paste, grep – regular expression – sort.

UNIT-II: (8 hrs)

Process Management: Process– Concept - Scheduling Algorithms - Short term and long term process scheduling policies – Scheduling Criteria - Multiple Processor Scheduling. **CPU Scheduling:** Scheduling Criteria – Scheduling Algorithms - **Process Synchronization and Deadlocks:** The Critical-section Problem – Petersons solution – Mutex locks - Semaphores – Monitors, Deadlock Prevention and Avoidance, Deadlock Detection and Deadlock Recovery. **Process Utilities:** sh process, Parents and children, Process status, System process, Mechanism of process creation, Internal and external commands, running jobs in background, KILL, NICE, Job control, at and batch, cron - Case Study on Processes in LINUX.

UNIT-III: (8 hrs)

File Organisation: Logical and Physical File structure – File Allocation methods, -Linked and Index Allocation - File Protection and Security - Directory structure - single level, two

level, tree structure - Free Space Management - Allocation Methods - Efficiency and Performance – Recovery – FAT32 and NTFS.**FileSystem:**File Access Permission – File Comparisons - View Files – Listing files with attributes – Wildcards - Translating Characters - Links and its types - The File System – Partitions, File Systems, Kernel Accesses – Mounting – umask, ulimit - I/O redirection – Pipes - Case Study on LINUX File System.

UNIT-IV : (8 hrs)

Memory Management: Techniques - Single Partition Allocation - Multiple Partition Allocation – Swapping - Paging and Segmentation - Segmented-Paged Memory Management Techniques - Logical and Physical Address space – Address Mapping - Demand paging - Virtual memory, protection and address mapping hardware, Page fault, Page replacement and Page removal algorithms. **Device Management:** Classification of device according to speed, Disk structure - Disk scheduling – FCFS scheduling, SSTF scheduling - Access method and storage capacity. **Disk Utilities:** Disk usage, disk free, - Backups- cpio, tar, System calls for file management, directory management - Case Study on Memory Management in LINUX.

UNIT-V: (8 hrs)

Security: Security Environment – Operating System Security – Controlling Access to resources – Formal models of Secure systems - Basics of cryptography – Authentication – Exploiting Software – Insider Attacks – Malware –Defenses. **Virtualization and the Cloud:** Requirements– Type 1 and Type 2 Hypervisors – Techniques for efficient virtualization – Memory virtualization – I/O Virtualization –Virtual Appliances – Virtual machines on multicore CPUs – **Clouds:** Clouds as service – Virtual machine migration – check pointing - Case Study on Security in LINUX.

Text Books:

1. Silberschatz, Abraham, Peter Baer Galvin and Greg Gagne. Operating System Concepts. 10th ed. Addison Wesley. (Units 1 to 4 - Chapters 1-4, 6-13)
2. Sumitabha Das. UNIX – Concepts & Applications. 3rd ed. New Delhi: TataMcGraw Hill, 2000.
3. Tanenbaum S., Andrew, Herbert Bos. Modern Operating Systems. 4th ed. Pearson (Unit 5 - Chapter 7, 9, Case Studies – Chapter 10)
4. YukunLiu,YongYue,Liwei Guo UNIX Operating System The Development Tutorial via UNIX Kernel Services. Beijing: Higher Education Press (Chapters 1,2, 6-10)

Reference Books:

1. Kanetkar Yashwant. UNIX Shell Programming. BPB.
2. Rosen Kenneth, Douglas Host, Rachel Klee and Richard Rosinski.UNIX: The Complete Reference. 2nd ed. McGraw Hill/Osborne, 2007.
3. Sobell M. G. A Practical Guide to Linux Commands, Editors, and Shell Programming. USA: Pearson Education

Web References

- www.tutorialspoint.com/unix
- www.unixtutorial.org/
- www.guru99.com/unix-linux-tutorial.html

Java Programming Lab
(Course code : 23PDSE13)

SEMESTER– I	HOURS – 2	CREDITS – 1

Objective:
To be able to implement/ apply the concepts of data structures and algorithms as well as operating systems in developing standalone and web applications.

Outcomes:
At the end of the course the students must be able to:
CO1: Associate the problems with data structures and operating system concepts. (K3)
CO2: Apply arrays/ built-in functions to solve complex problems. (K5)
CO3: Demonstrate passing values to functions associating with interrupts/ system calls. (K4)
CO4: Evaluate the performance of algorithms and concepts. (K5)
CO5: Solve problems using control structures and looping structures. (K6)
CO6: Develop software/ web applications by applying the concepts. (K6)

List of Practical's
I. Java
1. Problems using Control Structures
2. Problems using Looping Structures
3. Problems using Arrays
4. Problems using Constructors
5. Implement pass by value, pass by reference
6. Implement Linear Search and compute the time complexity
7. Problems on Inheritance using Interface
8. Implement sorting algorithms and compute the time complexity
9. Problems using Multithreading
10. Implement Array Operations using Java
11. Implement Singly Linked List
12. Implement Stack
13. Implement Queue
14. Implement Binary search and compute the time complexity.
15. Implement Tree Traversals
16. Solve N Queens problem

II. Web Programming Lab
1. Simple web page creation using HTML
2. HTML form validation using VB script
3. Simulating a Calculator using VB script
4. Testing Request and Response Objects
5. Testing Application and Session Objects
6. Testing Validation Controls
7. Database Access - ADO.NET
8. Use of Data Grid and Data List Viewer
9. Develop a college application (automate admission and selection process)
10. Develop a web site to handle student information (Attendance and Marks Details)

III. Operating System Lab
1. Basic Unix Commands
2. File Manipulation Commands
3. Working with Unix Editors
4. Program for System Calls of Unix Operating Systems

5. Working with Shell (Branching and Looping structures)
6. CPU Scheduling Algorithms (Priority/ Round Robin/ FCFS/ SJF)
7. Problems using Semaphores
8. IPC using Shared Memory
9. Algorithm for Deadlock Detection
10. Algorithm for Deadlock Avoidance

Text Books

1. Cay Horstmann, Big Java, Wiley India Edition, 2nd edition, 2016.
2. Mridula Parihar, "ASP.NET Bible", Published by Hungry Minds, 2002, Third Avenue New York, NY 10022, ISBN: 0-7645-4816-6.
3. Maurice J. Bach, The Design of UnixOperating System, Prentice-Hall International Inc., USA.

Reference Books

- E. Balaguruswamy, Programming with Java A Primer, Tata McGraw Hill, 3rd Ed, 2006.
- 2.Stephanie B, Dale G, Eric J, The J2EE tutorial, AddisonWesley,New York, 2002.
- Bill Evjenet, Professional ASP.NET 2.0, Special Edition, Wrox Press, 2006,ISBN:9780470041789.
- Stephen Walther, ASP.NET 2.0 Unleashed, Sams Publications, 2006,ISBN: 9780672328237.
- Ken Rosen, James Farber, Douglas Host, Rachel Klee, Richard Rosinski, UNIX - The Complete Reference, 2nd Edition, Mc Graw Hill, 2007.ISBN:9780072263367.

PYTHON PROGRAMMING
(Course code: 23PDSC21)

SEMESTER– II	HOURS – 6	CREDITS – 5

Objective:
To be able to think logically and develop interactive programs using the python constructs, functions, data structures, classes and objects, files.

Outcomes:
At the end of the course the students will be able to:
CO1: Explain the components of a computer, demonstrate the appropriate use of data types, mathematical functions, and strings in a program. (K2)
CO2: Describe the use of selection and looping constructs, compare, and choose an appropriate construct for a given problem. (K2)
CO3: Applyfunctions, classes, and objects to examine their use in solving real time problems. (K3)
CO4: Illustrate the use of strings and lists in designing solutions. (K5)
CO5: Apply Tuples, sets, dictionaries, and files, compare programs with and without files, develop applications using the different data structures. (K4)
CO6: Develop applications and evaluate program using Numpy, Pandas, Matplotlib and SciKit. (K6)

UNIT-I: (18 hrs)
Introduction to Computers, Programs and Python: History–Programming Style and Documentation –Errors –Graphics Programming. **Elementary Programming**– Evaluating Expressions and Operator Precedence – Augmented Assignment Operators – Type Conversion and Rounding. **Mathematical Functions, Strings and Objects** –Common Python Functions – Strings and Characters –Objects and Methods – Formatting Numbers and Strings – Drawing shapes. **Selections** – Boolean Types, Values and Expressions – Generating Random Numbers – Loops.

UNIT-II: (18 hrs)
Functions –Return single and multiple values – Passing Arguments–Abstraction and Stepwise Refinement – Recursion. **Objects and Classes** –UML Class Diagrams – Immutable vs Mutable Objects – Hiding Data Fields – Class Abstraction and Encapsulation. Inheritance and Polymorphism–**Strings and Special Methods** –Str class – Operator Overloading and Special Methods.

UNIT-III: (18 hrs)
Lists –Copying Lists – Passing Lists to Functions – Returning a List from a Function – Searching, Sorting Lists. **Multidimensional Lists** –Two Dimensional Lists – Multidimensional Lists. **Tuples, Sets and Dictionaries** –Tuples – Sets – Comparing the Performance of Sets and Lists – Dictionaries. **Files and Exception Handling** – Text Input and Output –Retrieving Data from Web – Exception Handling – Raising Exceptions – Processing Exceptions using Exception Objects – Defining Custom Exception Classes – Binary IO Using Pickling.

UNIT-IV: (18 hrs)
Advanced Python: Database - Connections - Executing queries - Transactions. **GUI Programming:** Tkinter Programming. Networking: Protocols, server-client architecture, TCP/IP and UDP communication. **Numpy:** Introduction - Basics of NumPy Array – Computation on NumPy Array – Aggregations – Broadcasting – Comparisons, Masks and Boolean Logic – Sorting Arrays – NumPy Structured Array.

UNIT-V: **(18 hrs)**

Data Manipulation: Panda Objects – Data Indexing and Selection - Operating Data on Pandas – Handling Missing Data – Hierarchical Indexing –Combining Data Sets – Vectorized String Operations – Working with Time Series- Statistics with Pandas – Data Frames. **Visualization**: Matplotlib - Scatter plot - Line plot - Bar chart - Histogram - Box plot. Seaborn for Statical plots, Interactive Dynamic visualizations, SciKit for Machine learning.

Text Books:

1. Y. Daniel Lang, Introduction to Programming using Python, 2nd Edition, Pearson Education Inc., 2013.
2. **Unit I:** Ch.1, Ch.2, Ch.3, Ch.4, Ch.5.**Unit II:** Ch.6, Ch.15(15.1,15.2,15.4), Ch.7, Ch.12, Ch.8. **Unit III:** Ch. 10, Ch.11, Ch. 13, Ch. 14. Unit IV: Ch.9.
3. Jeeva Jose and P. Sojan Lal, "Introduction to Computing and Problem Solving with PYTHON", Khanna Book Publishing Co. (P) Ltd., 2016.
4. **Unit IV: Ch. 12**
4. Jake Vander Plas, "Python Data Science Handbook: Essential Tools for Working withData", 1st Edition, O'Reilly Media, 2016.
5. **Unit IV: Ch.2, Unit V: Ch. 3, Ch. 4**

References Books:

1. Allen B. Downey. Think Python. How to Think Like a Computer Scientist, 2ndedition, O'Reilly Publishers, 2016.
2. Corey Wade, et al:The Python Workshop, 2nd Edition, Packt, 2022.
3. David Beazley, Brian K. Jones. Python Cookbook: Recipes for Mastering Python 3, 3rd Edition, 2013 Harsh Bhasin. Python for Beginners. New Age International Publishers,2018.
4. Martin C. Brown. Python: The Complete Reference. McGraw Hill Education; Fourth edition, 2018.

Web References:

- https://www.geeksforgeeks.org/web-technology
- https://aclm.in/advanced-web-technology-training.html
- https://www.bcit.ca/programs/advanced-web-technologies
- https://docs.python.org/3/tutorial/index.html
- https://www.python-course.eu/advanced_topics.php

PYHTON PROGRAMMING LAB
(Course code: 23PDSC22)

SEMESTER– II	HOURS – 6	CREDITS – 5

Objective:
To be able to apply appropriately the python programming knowledge gained and develop computer-based solutions for a given problem.

Outcomes:
At the end of the course the students will be able to:
CO1: Explain the components of a computer, demonstrate the appropriate use of data types, mathematical functions, and strings in a program. (K2)
CO2: Describe the use of selection and looping constructs, compare, and choose an appropriate construct for a given problem. (K2)
CO3: Apply functions, classes, and objects to examine their use in solving real time problems. (K3)
CO4: Illustrate the use of strings and lists in designing solutions. (K5)
CO5: Apply Tuples, sets, dictionaries, and files, compare programs with and without files, develop applications using the different data structures. (K4)
CO6: Develop applications and evaluate program using Numpy, Pandas, Matplotlib and SciKit. (K6)

UNIT-I: (18 hrs)
Installation of the required software - Programs using basic data types and operators - Programs involving Mathematical functions - Program in String Manipulations - Programs using different forms of if statement.

UNIT-II: (18 hrs)
Drawing various shapes using turtle - Programs involving repeated execution of a set of statements - Programs using break and continue - Programs using random - Modular programming using functions - Programs using positional - Default argument.

UNIT-III: (18 hrs)
Programs using pass by value, pass by reference - Programs using classes and objects - Programs using Inheritance - Programs on Str class and special methods - Programs using Lists and List manipulation - Programs using Two-Dimensional Lists.

UNIT-IV: (18 hrs)
Programs using Tuple and its methods - Programs with Set and Set manipulation - Programs using Dictionaries – Program comparing the performance of Sets and Lists - Programs handling Text Files - Programs handling Binary Files - Programs handling exceptions.

UNIT-V: (18 hrs)
Data Analysis - NumPy Arrays - Sorting and Searching on Arrays. **Pandas:** Data Series - Data Frame - Combining and Merging Data Sets - Handling Missing Values – Filter - Grouping and Aggregation. **Visualization:** MatplotLib– Line Chart, Scatter Plot, Histogram. **Seaborn**: Boxplot, HeatMap.

Text Book
1. Y. Daniel Lang, Introduction to Programming using Python, 2nd Edition, Pearson Education Inc., 2013.
2. Jake Vander Plas, "Python Data Science Handbook: Essential Tools for Working with Data", 1st Edition, O'Reilly Media, 2016

Reference Books:

1. Allen B. Downey. Think Python. How to Think Like a Computer Scientist, 2nd edition, O'Reilly Publishers, 2016.
2. Corey Wade, et al:*The Python Workshop*, 2nd Edition, Packt, 2022.
3. David Beazley, Brian K. Jones. Python Cookbook: Recipes for Mastering Python 3, 3rd Edition, 2013 Harsh Bhasin. Python for Beginners. New Age International Publishers,2018.
4. Martin C. Brown. Python: The Complete Reference. McGraw Hill Education; Fourth edition, 2018.

Web References:

- https://realpython.com
- http://docs.python.org
- http://diveintopython.org
- https://www.w3schools.com/python/
- https://www.tutorialspoint.com/python/index.htm

STATISTICS - II
(Course code : 23PDSC23)

SEMESTER– II	HOURS – 6	CREDITS – 4

Objective:
To develop knowledge and understand fundamental concepts in probability and statistics.

Outcomes:
At the end of the course the students will be able to:
CO1: Identify different steps involved in hypothesis testing. (K1)
CO2: Restate the applied principles of statistics. (K2)
CO3: Apply knowledge and skills in theoretical, computational and application- oriented statistics (K3)
CO4: Analyze the methods of using variancein statistical analysis (K4)
CO5: Evaluate the concept of non-parametric tests (K5)
CO6: Design solutions for problem using Hypothesis testing (K6)

UNIT-I: (18 hrs)
INTRODUCTION: Population and Statistics – Finite and Infinite population – Parameter and Statistics – Types of sampling – Sampling Distribution – Sampling Error – Standard Error – Test of significance –concept of hypothesis – types of hypotheses – Errors in hypothesis-testing – Critical region – level of significance – Power of the test – p-value. **Hypothesis testing:** Introduction-Significance Levels-Tests Concerning the Mean of a Normal Population-Case of Known Variance-Case of Unknown Variance: The t-Test-Testing the Equality of Means of Two Normal Populations-Case of Known Variances-Case of Unknown Variances-Case of Unknown and Unequal Variances-The Paired t-Test-Hypothesis Tests Concerning the Variance of a Normal Population-Testing for the Equality of Variances of Two Normal Populations-Hypothesis Tests in Bernoulli Populations-Testing the Equality of Parameters in Two Bernoulli Populations-Tests Concerning the Mean of a Poisson Distribution-Testing the Relationship Between Two Poisson Parameters.

UNIT-II: (12 hrs)
Hypothesis Testing-II: Students t-distribution and its properties (without proofs) – Single sample mean test – Independent sample mean test – Paired sample mean test – Tests of proportion (based on t distribution) – F distribution and its properties (without proofs) – Tests of equality of two variances using F-test – Chi-square distribution and its properties (without proofs) – chisquare test for independence of attributes – Chi-square test for goodness of fit.

UNIT-III: (12 hrs)
Regression: Introduction-Least Squares Estimators of the Regression Parameters-Distribution of the Estimators-Statistical Inferences About the Regression Parameters-Inferences Concerning β -Inferences Concerning α- Inferences Concerning the Mean Response $\alpha+\beta$ x 0 – Prediction Interval of a Future Response-Summary of Distributional Results- The Coefficient of Determination and the Sample Correlation Coefficient-Analysis of Residuals: Assessing the Model-Transforming to Linearity- Weighted Least squares-Polynomial Regression – Multiple Linear Regression-Predicting Future Responses – Logistic Regression Models for Binary Output Data.

UNIT-IV: (12 hrs)
Analysis of variance: Introduction-An Overview-One-Way Analysis of Variance-Multiple Comparisons of Sample Means-One-Way Analysis of Variance with Unequal Sample Sizes-Two-Factor Analysis of Variance: Introduction and Parameter Estimation-Two-Factor Analysis of Variance: Testing Hypotheses-Two-Way Analysis of Variance with Interaction.

Goodness of fit tests and categorical data analysis: Introduction-Goodness of Fit Tests When All Parameters Are Specified-Determining the Critical Region by Simulation-Goodness of Fit Tests When Some Parameters Are Unspecified-Tests of Independence in Contingency Tables -Tests of Independence in Contingency Tables Having Fixed Marginal Totals-The Kolmogorov–Smirnov Goodness of Fit Test for Continuous Data.

UNIT-V: (12 hrs)

Nonparametric hypothesis tests: Introduction-The Sign Test-The Signed Rank Test-The Two-Sample Problem-The Classical Approximation and Simulation-Wilcoxon Signed Rank Test for one and paired samples-The Runs Test for Randomness -Median test and Mann-Whitney-Wilcoxon tests for two samples.

Text Books:

1. Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020.**(Unit I, Unit II, Unit III, Unit V)**
2. Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists,Elsevier Academic Press, UK, Fifth Edition, 2023.**(Unit IV)**
3. Brian Caffo, Statistical Inference for Data Science, Learn pub, 2016.**(Unit I)**

Reference Books:

1. Allen B. Downey,Think Stats- Exploratory data analysis, O'Reilly, 2nd Edition
2. Erwin Kreyszig, Advanced Engineering Mathematics, Wiley Publications, Tenth Edition
3. Jim Frost, Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries.

Web References:

- https://onlinestatbook.com/2/
- https://www.simplilearn.com/tutorials/statistics-tutorial
- https://towardsdatascience.com/fundamentals-of-statistics-for-data-scientists-and-data-analysts-69d93a05aae7

SOFTWARE ENGINEERING FOR DATA SCIENCE
(Course Code : 23PDSE21)

SEMESTER– II	HOURS – 4	CREDITS – 3

Objective:
To understand the software engineering principles and ensure software quality**.**

Outcomes:
At the end of the course the students will be able to:
CO1: Describe the Software Engineering Principles (K2)
CO2: Apply Software Life Cycle Models for Software Development (K2)
CO3: Examine Requirements Engineering skills and gather Requirements (K3)
CO4: Analyze the quality of a Software (K4)
CO5: Select appropriate testing methodologies and tools (K5)
CO6: Create a simple application using SDLC method. (K6)

UNIT-I: (12 hrs)
Software and Software Engineering: The nature of software – Software Engineering – The Software Process – Software Engineering Practice – Software Myths. **Process Models:** A Generic Process Model – Process Assessment and Improvement – Prescriptive Process Models – Product and Process. **Agile Development:** Introduction – Agility and Cost of Change – Agile Process – Scrum – Other Agile Frameworks.

UNIT-II: (10 hrs)
Recommended Process Model: Requirements Definition – Preliminary Architectural Design – Resource Estimation – First Prototype Construction – Prototype Evaluation – Prototype Evolution – Prototype Release – Maintain Release Software. **Human Aspects of Software Engineering:** Characteristics of a Software Engineer –Psychology of Software Engineer – Software Team – Team Structures –Impact of social media– Global Teams. **Principles that guide practice**: Core Principles – Principles that guide each Framework Activity – Communication Principles – Planning Principles – Modeling Principles – Construction Principles – Deployment Principles.

UNIT-III: (11 hrs)
Understanding Requirements: Requirements Engineering – Establishing the groundwork – Requirements Gathering – Developing Use Cases -Building the Analysis Model – Negotiating Requirements – Requirements Monitoring – Validating Requirements. **Requirements Modeling – A Recommended Approach:** Requirements Analysis – Scenario-Based Modeling – Class-Based Modeling – Functional Modeling –Behavioral Modeling.

UNIT-IV: (11 hrs)
Design Concepts: Design within the context of Software Engineering – The Design Process – Design Concepts – The Design Model. **Quality and Security**: Introduction – Software Quality – The Software Quality Dilemma – Achieving Software Quality. **Software Quality Assurance:** Background Issues – Elements of Software Quality Assurance – SQA Process and Product Characteristics – SQA Tasks, Goals and Metrics – Formal Approaches – Statistical SQA – Software Reliability – ISO 9000 Quality standards – SQA Plan

UNIT-V: (10 hrs)
Software Testing -Component Level: A Strategic Approach to Software Testing – Planning and Record Keeping– Test-Case Design – White-box Testing – Black-Box Testing – Object-oriented Testing. **Software Testing – Integration Level:** Software Testing Fundamentals –

Integration Testing – Artificial Intelligence and Regression Testing – Integration Testing in the OO context – Validation Testing – Testing Patterns

Text Books:

1. Pressman, Roger S., and Bruce R. Maxim. Software Engineering: A Practitioner's Approach, Ninth Edition, 2020.
   Unit I: Ch. 1, Ch. 2, Ch. 3, **Unit II:** Ch. 4, Ch. 5, Ch. 6, **Unit III:** Ch. 7, Ch. 8,
   Unit IV: Ch. 9, Ch. 15, Ch. 17, **Unit V:** Ch. 19, Ch. 20,

Reference Books:

1. Martin, Robert C. Agile software development: principles, patterns, and practices. Prentice Hall, 2002.
2. Schach, Stephen R. Object-oriented software engineering. McGraw-Hill, 2008.
3. Sommerville, Ian. "Software engineering 9th Edition." ISBN-10 137035152, 2011.

Web References:

1. https://www.d.umn.edu/~gshute/27often/principles.html

DISTRIBUTED SYSTEMS
(Course code : 23PDSE21)

SEMESTER– II	HOURS –4	CREDITS – 3

Objective:
To learn the principles, architectures, Processes, Communication, Co-ordination, consistency and Replication in Distributed Systems.

Outcomes:
At the end of the course the students will be able to:
CO1: Describe the significance of Distributed Systems(K2)
CO2: Explain the architecture of Distributed Systems (K2)
CO3: Apply the different types of processes/ roles in Distributed Systems (K3)
CO4: Classify the rules the communicating processes must adhere to (K4)
CO5: Compare the issues in Distributed Systems (K5)
CO6: Build the systems using synchronization techniques. (K6)

UNIT-I: **(10 hrs)**
Introduction –Introduction to Distributed Systems – Design Goals – Types of Distributed Systems. **Architectures –** Architectural Styles – Middleware Organization – System Architecture – Example Architectures.

UNIT-II: **(10 hrs)**
Processes: Threads – Virtualization – Clients – Servers – Code Migration. **Communications:** Foundations – Remote Procedure Call -Basic RPC operation, Parameter Passing, RPC based Application Support.

UNIT-III: **(10 hrs)**
Communications: Message Oriented Communication – Simple transient Messaging with Sockets, Advanced Transient Messaging, Message Oriented Persistent Communication – Multicast Communication.

UNIT-IV: **(10 hrs)**
Naming: Names, Identifiers and Addresses – Flat naming – Structured naming – Attribute-based naming. **Co-ordination:** Clock Synchronization – Logical Clocks – Mutual Exclusion – Election Algorithms – Distributed Event Management.

UNIT-V: **(12 hrs)**
Consistency and Replication: Introduction – Data-centric Consistency Models – Client-Centric Consistency Models – Replica Management. **Fault Tolerance:** Introduction.

Text Book:
1. Andrew S. Tannenbaum and Maarten Van Steen, "Distributed Systems: Principles and Paradigms", Third Edition, Pearson, 2017.
   Unit I: Ch. 1, **Unit II:** Ch. 2, **Unit III:** Ch. 3, **Unit IV:** Ch. 4, Ch. 5
   Unit V: Ch. 6 – 6.1,6.2,6.3,6.4,6.6, Ch. 7-7.1 to 7.4, Ch. 8-8.1

Reference Books:
1. Andrew S. Tannenbaum and Maarten Van Steen, "Distributed Systems: Principles and Paradigms", Third Edition, Pearson, 2017.

Web References:
1. https://www.tutorialspoint.com/Distributed-Systems
2. https://link.springer.com/article/10.1007/s00607-016-0508-7

INFORMATION SECURITY AND ETHICS
(Course code : 23PDSE21)

SEMESTER– II	HOURS –4	CREDITS – 3

Objective:
To introduce and familiarize the students to security issues in computing, core concepts and vocabulary of computer security.

Outcomes:
At the end of the course the students will be able to:
CO1: Associate all aspects of computer security, including users, software, devices, operating systems, networks, law, and ethics. (K2)
CO2: Apply cryptography an essential tool that is critical to computer security. (K2)
CO3: Analyze the different aspects of computer security and privacy. (K3)
CO4: Evaluate the aspects of computer security. (K4)
CO5: Develop a system that uses user authentication, prevents malicious code execution, encrypts the data, protects privacy, implements firewall, detects intrusion, and more. (K5)
CO6: Build the systems embedded with synchronization and security enforcements techniques. (K6)

UNIT-I: (10 hrs)
Security Problem in Computing: Meaning of "Secure" – Attacks - Meaning of Computer and information Security - Computer Criminals - Methods of Defense. **Cryptography:** Terminology and Background - Principles of Cryptography - Cryptography tools - Substitution Ciphers - Transpositions (Permutations) – Making "Good" Encryption Algorithms - The Data Encryption Standard (DES) – AES Encryption Algorithm - Public Key Encryption - The Uses of Encryption - Digital Signatures and Certificates - Hybrid Cryptography Systems - Steganography - Protocols for secure communication.

UNIT-II: (10 hrs)
Program Security: Secure Programs – Non-malicious Program Errors - Viruses and Other Malicious Code - Targeted Malicious Code - Controls against Program Threats. **Security Issues in Social Networking:** Acceptable Use Policies - Reasons for social media being hazardous to the corporate network - Balancing Security and Social Networking in business - Precautions that can be taken to secure the private information.

UNIT-III: (10 hrs)
Database and Data Mining Security: Introduction to Databases - Security Requirements - Reliability and Integrity – Sensitive Data - Inference - Multilevel Databases - Proposals for Multilevel Security – Data Mining. **Security in Networks:** Network Concepts - Threats in Networks - Network Security Controls - Firewalls – Intrusion Detection Systems - Secure E-Mail.

UNIT-IV: (10 hrs)
Administering Security: Security Planning - Risk Analysis - Organisational Security Policies - Physical Security. **Economics of Cyber security:** Making a Business Case - Quantifying Security - Modeling Cyber security.

UNIT-V: (10 hrs)
Privacy in Computing: Privacy Concepts - Privacy Principles and Policies - Authentication and Privacy – Data Mining - Privacy on The Web - E-Mail Security - Impacts on Emerging Technologies. **Legal and Ethical Issues in Computer Security:** Protecting Programs and

Data - Information and the Law - Rights of Employees and Employers - Redress for Software Failures - Computer Crime - Ethical Issues in Computer Security - Case Studies of Ethics.

Text Book:

1. Pfleeger Charles P, Shari Lawrence Pfleeger. Security in Computing, Released January 2015, Pearson, ISBN: 9780134085074.

Reference Books:

1. Bahadur, Gary. Securing the Clicks Network Security in the Age of social media. 1st ed. McGraw-Hill, 2012.
2. Daswani, Neil, Christoph Kern and Anita Kesavan. Foundations of Security: What Every Programming Needs to Know. Apress, 2007.

Web References:

- http://www.trendmicro.fr/media/wp/securityguide-social-networks-whitepaper-en.pdf
- http://paper.ijcsns.org/07_book/201306/20130619.pdf

DISCRETE MATHEMATICS
(Course code : 23PDSE22)

SEMESTER– II	HOURS –4	CREDITS –3

Objective:
To develop knowledge and understand concepts of mathematical induction, logic, functions and relations.

Outcomes:
At the end of the course the students will be able to:
CO1: Recall the equivalence of statements using mathematical logic (K1)
CO2: Restate inference theory and predicate calculus to understand partial order and partition. (K2)
CO3: Apply the fundamental principles of Combinatorial Counting techniques (K3)
CO4: Analyse generating functions and their utility in solving recurrence relations (K4)
CO5: Assess graph models and tree structures with basics and significance of traversability. (K5)
CO6: Develop an algorithm using Tree concepts. (K6)

UNIT-I: (12 hrs)
Sets, Sequences and Functions: Sets-Some Special Sets-Set Operations-Functions-Sequences-Properties of Functions -Propositions-Conditional Propositions and Logical Equivalence-Arguments and Rules of Inference-Quantifiers-Nested Quantifiers. **Elementary Logic:** Informal Introduction-Propositional Calculus-Getting Started with Proofs-Methods of Proof-Logic in Proofs-Analysis of Arguments.

UNIT-II: (12 hrs)
Relations: Relations-Digraphs and Graphs-Matrices-Equivalence Relations and Partitions-The Division Algorithm and Integers Mod p. **Induction and Recursion:** Loop Invariants-Mathematical Induction-Big-Oh Notation-Recursive Definitions-Recurrence Relations-More Induction- Euclidean Algorithm

UNIT-III: (12 hrs)
Counting: Basic Counting Techniques–Elementary Probability – Inclusion–Exclusion and Binomial Methods–Counting and Partitions–Permutations and Combinations, Binomial Coefficients and Identities, Equivalence Relations, Generalized Permutations and Combinations, Generating Functions, Inclusion–Exclusion, Applications of Inclusion – Exclusion – Pigeon–Hole Principle. **Algorithms:** Introduction–Examples of Algorithms–Analysis of Algorithms–Recursive Algorithms.

UNIT-IV: (12 hrs)
Graphs: Graphs–Paths and Cycles–Edge Traversal Problems-Hamiltonian Cycles and the Traveling Salesperson Problem–A Shortest–Path Algorithm-Representations of Graphs-Isomorphisms of Graphs–Planar Graphs. **Trees:** Trees-Terminology and Characterizations of Trees-Rooted Trees-Vertex Traversal Problems-Spanning Trees-Minimal Spanning Trees-Binary Trees- Tree Traversals-Decision Trees and the Minimum Time for Sorting – Isomorphism of Trees.

UNIT-V: (12 hrs)
Recursion and Digraphs: General Recursion-Depth-First Search Algorithms-Polish Notation-Weighted Trees-Digraphs-Digraphs Revisited-Weighted Digraphs and Scheduling Networks-Digraph Algorithms.

Text Books:

1. Kenneth A. Ross and Charles R. B. Wright, Discrete Mathematics, Pearson Education, 5th Edition.
2. Richard Johnsonbaugh, Discrete Mathematics, Pearson Education, 8th Edition, 2018

Reference Books:

1. Discrete Mathematics and its Applications (6th edition), Kenneth H. Rosen, Tata McGraw Hill, Bombay, India.
2. Discrete Mathematics with Applications Susanna S. Epp, Brooks/Cole 2011.
3. Discrete Mathematics an Introduction to Proofs and Combinatorics, Kevin Ferland, Houghton Mifflin Company, 2009.

Web References:

- https://www.tutorialspoint.com/discrete_mathematics/discrete_mathematics_introduction.htm
- https://www.cs.odu.edu/~toida/nerzic/content/intro2discrete/intro2discrete.html

APPLIED PROBABILITY
(Course code : 23PDSE22)

SEMESTER– II	HOURS –4	CREDITS –3

Objective:
To develop knowledge and understand fundamental concepts and applications of probability.

Outcomes:
At the end of the course the students will be able to
CO1: Recite the principal concepts about probability. (K1)
CO2: Summarize combinatorics and convexity (K2)
CO3: Examine the nature and properties of density functions and hence determine the moments and moment generating functions of any random variable (K3)
CO4: Illustrate the value of point estimators using the method of moments and method of maximum likelihood (K4)
CO5: Assess the illustration of data through advanced chart types. (K5)
CO6: Formulate discrete-time and continuous-time Markov chains (K6)

UNIT-I: (10 hrs)
Basic Notions of Probability Theory: Introduction-Probability and Expectation-Sample Spaces and Events-Random Experiments, Sample Spaces, Events, Counting Techniques-Interpretations and Axioms of Probability-Addition Rules-Conditional Probability-Multiplication and Total Probability Rules-Independence-Bayes' Theorem-Random Variables-Distributions, Densities, and Moments-Convolution-Random Vectors-Multivariate Normal Random Vectors. **Calculation of Expectations:** Introduction-Indicator Random Variables and Symmetry-Conditioning-Moment Transforms-Tail Probability Methods-Moments of Reciprocals and Ratios-Reduction of Degree-Spherical Surface Measure.

UNIT-II: (10 hrs)
Convexity and Combinatorics: Introduction-Convex Functions-Minimization of Convex Functions-The MM Algorithm-Moment Inequalities-Combinatorics-Introduction-Bijections-Inclusion-Exclusion -Applications to Order Statistics-Catalan Numbers-Pigeonhole Principle-Combinatorial Optimization-Introduction-Quick Sort-Data Compression and Huffman Coding-Graph Coloring.

UNIT-III: (10 hrs)
3.1 Discrete Random Variables: Probability Distributions and Probability Mass Functions–Cumulative Distribution Functions–Mean and Variance of a Discrete Random Variable–Discrete Uniform Distribution–Binomial Distribution–Geometric and Negative Binomial Distributions–Hypergeometric Distribution–Poisson Distribution.**3.2 Continuous Random Variables:** Probability Distributions and Probability Density Functions–Cumulative Distribution Functions–Mean and Variance of a Continuous Random Variable–Continuous Uniform Distribution–Normal Distribution – Normal Approximation to the Binomial and Poisson Distributions–Exponential Distribution-Erlang and Gamma Distributions-Weibull Distribution-Lognormal Distribution–Beta Distribution.

UNIT-IV: (10 hrs)
4.1 Two or More Random Variables: Joint Probability Distributions-Marginal Probability Distributions-Conditional Probability Distributions-Independence-More Than Two Random Variables- Covariance and Correlation-Common Joint Distributions-Multinomial Distribution-Bivariate Normal Distribution-Linear Functions of Random Variables-General Functions of Random Variables. **4.2 Sampling Distributions and Point Estimation of Parameters:** Point Estimation-Sampling Distributions and the Central Limit Theorem-

General Concepts of Point Estimation-Unbiased Estimators-Variance of a Point Estimator - Standard Error: Reporting a Point Estimate-Mean Squared Error of an Estimator-Methods of Point Estimation-Method of Moments-Method of Maximum Likelihood-Bayesian Estimation of Parameters.

UNIT-V: (12 hrs)

5.1 Discrete-Time Markov Chains: Introduction-Definitions and Elementary Theory-Examples-Coupling-Convergence Rates for Reversible Chains-Hitting Probabilities and Hitting Times-Markov Chain Monte Carlo-simulated annealing.**5.2 Continuous-Time Markov Chains:** Introduction-Finite-Time Transition Probabilities-Derivation of the Backward Equations-Equilibrium Distributions and Reversibility-Examples-Calculation of Matrix Exponentials-Kendall's Birth-Death-Immigration Process.

Text Books:

1.Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for Engineers, 5thEdition, John Wiley & Sons, Inc.

(UNIT-I Ch2, UNIT-III Ch3,Ch4&UNIT-IV Ch5 &Ch7)

2.Stephan Wagner,Combinatorics,July 2011

(UNIT-II Ch-2,Ch-3,Ch-6)

3.Michael Mitzenmacher Eli Upfal, Probability and Computing Randomized Algorithms and Probabilistic Analysis, Cambridge University press, 2005.

(UNIT-V Ch7.1,Ch11.1,Ch11.4,Ch12.2)

Reference Books:

1. Mario Lefebvre, Applied Probability and Statistics, Springer Newyork, 2006.
2. Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, Elsevier Academic Press, UK, Fifth Edition, 2023

Web References:

- https://open.umn.edu/opentextbooks/textbooks/256
- https://www.intechopen.com/books/12021

(Course code : 23PDSE22)

SEMESTER– II	HOURS –4	CREDITS –3

Objective:
To study of model formulation and apply the mathematical results and numerical techniques of optimization theory to real world problems.

Outcomes:
At the end of the course the students will be able to:
CO1: Explain the fundamentals concepts of Linear Programming (K1)
CO2: Use classical optimization techniques and numerical methods of optimization. (K2)
CO3: Enumerate fundamentals of Integer programming technique and apply different techniques to solve various optimization problems (K3)
CO4: Describe the basics of different Heuristic algorithms and solve dynamic programming problems. (K4)
CO5: Defend queuing systems and understand constrained and unconstrained problems(K5)
CO6: Formulate discrete-time and continuous-time Markov chains (K6)

UNIT-I: (10 hrs)
Modelling with Linear programming: Two variable LP model – Graphical LP solution – Applications. **Simplex method and sensitivity analysis:** Simplex method- Artificial starting solution - Special cases in simplex method- Graphical sensitivity analysis.

UNIT-II: (10 hrs)
Duality and post-optimal Analysis: Definition of Dual problem - Primal-Dual Relationships-Additional Simplex algorithms- Post optimal analysis. **Advanced Linear Programming:** Simplex method fundamentals-Revised Simplex Method, Bounded-Variable Algorithm, Duality, Parametric programming.

UNIT-III: (10 hrs)
Goal Programming: Goal programming formulation - Goal Programming algorithms. **Integer Programming:** Formulation and Applications-Cutting Plane Algorithm-Branch and Bound Method.

UNIT-IV: (10 hrs)
Heuristic Programming: Greedy Heuristics- Meta heuristic - Tabu Search algorithm - Constraint programming. **Deterministic dynamic programming:** Recursive nature of Dynamic programming computations - Forward and backward recursion- Selected DP applications - Knapsack/Fly-away kit/cargo-loading model- Investment Models-Inventory models.

UNIT-V: (12 hrs)
Queuing Systems: Pure birth and Pure death models- Generalized Poisson queuing model, single server models. **Classical optimization theory:** Unconstrained problems - Constrained problems.

Text Books:
1.Hamdy A. Taha, Operations Research- An Introduction, 10th Edition, Pearson Education – 2017.

Reference Books:
1. L. R. Foulds, Optimization Techniques, Springer Utm, 1981.
2. Garrido José M. Introduction to Computational Models with Python. CRC Press, 2016.

Web References:
- https://www.pre-scient.com/knowledge-center/optimization-problems/optimization-problems.html
- https://www.shsu.edu/~eco_dgf/web_chapter_a.pdf

EMERGING TECHNOLOGIES IN DATA SCIENCE
(Course code : 23PDSS21)

SEMESTER– II	HOURS – 4	CREDITS – 2

Objective:

To equip students with the knowledge and skills to apply data mining techniques using Weka and Orange for data preprocessing, classification, clustering, and advanced data mining tasks.

Outcomes:

At the end of the course the students will be able to:

CO1: Classify the fundamental concepts and principles of data mining. (K4)
CO2: Apply data preprocessing techniques and feature selection methods using Weka and Orange. (K2)
CO3: Implement classification and prediction models and evaluate their performance using Weka and Orange. (K3)
CO4: Apply clustering algorithms and association rule mining techniques using Weka and Orange. (K4)
CO5: Explore advanced data mining techniques, such as text mining and social network analysis, using Weka and Orange. (K5)
CO6: Interpret and visualize data mining results for decision-making and knowledge discovery. (K6)

Unit I: (12 Hrs)

Introduction to Data Mining: Overview of data mining concepts, process, and applications - Understanding data preprocessing and cleaning techniques - Introduction to Weka and Orange software interfaces and functionalities - Exploring data mining algorithms and techniques.

Unit II: (12 Hrs)

Data Preprocessing and Feature Selection: Data cleaning, transformation, and normalization techniques - Handling missing data and outliers - Feature selection and dimensionality reduction methods - Implementing data preprocessing techniques using Weka and Orange.

Unit III: (12 Hrs)

Classification and Prediction: Understanding classification algorithms, such as decision trees, naive Bayes, and k-nearest neighbors - Evaluating classification models and performance metrics - Utilizing ensemble methods for classification - Implementing classification and prediction models using Weka and Orange.

Unit IV: (12 Hrs)

Clustering and Association Rule Mining: Exploring clustering algorithms, including k-means and hierarchical clustering - Understanding association rule mining and frequent item set generation - Evaluating clustering and association rule models - Implementing clustering and association rule mining techniques using Weka and Orange.

Unit V: (12 Hrs)

Advanced Data Mining Techniques: Introduction to advanced techniques such as text mining and social network analysis - Applying sentiment analysis and text classification using Weka and Orange - Exploring integration and ensemble methods for data mining - Utilizing visualization tools for data mining results interpretation.

Text Books:

1. Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 4th edition 2016.

Reference Books:

1. "Weka 3 Data Mining with Java Cookbook" by Sargunan Sockalingam, Manoj Kumar, and Saman Halgamuge.
2. "Data Mining for the Masses with Orange: A Comprehensive Guide to Data Mining and Data Science in Orange" by Janusz Wojtusiak and Przemysław Biecek.
3. "Data Mining: Introductory and Advanced Topics" by Margaret H. Dunham.
4. "Introduction to Data Mining with Case Studies" by G.K. Gupta.

MACHINE LEARNING
(Course code : 23PDSC31)

SEMESTER– III	HOURS – 6	CREDITS – 5

Objective:
To understand the different types, steps and algorithms involved in Machine Learning Process

Outcomes:
At the end of the course the students will be able to:
CO1: Describe the data, essential steps for creating a typical ML model and the fundamentals of pattern classification. (K1)
CO2: Apply different ML algorithms using raw data and features. (K4)
CO3: Implement the essential techniques to reduce the number of features in a dataset and test the performance of predictive models. (K3)
CO4: Select multiple algorithms, combine, and produce ensembles, discuss the essential techniques for modeling linear relations. (K5)
CO5: Discuss the clustering algorithms, develop a Web application embedding a ML model. (K2)
CO6: Evaluate machine learning models using modern tools. (K5)

UNIT I: **(14 Hrs)**
Data Analytics with pandas and NumPy: NumPy and basic stats - Matrices - pandas library - Working with data - Null Values - Creating statistical graphs. **Giving Computers the ability to learn from data -** Introduction - Building intelligent systems to transform data into knowledge - The three different types of Machine Learning (ML) - Introduction to basic terminology and notations - A roadmap for building ML systems - Using Python for ML. **Training Simple ML Algorithms for Classification** - Early History of ML - Implementing a Perceptron learning algorithm - Adaptive linear neurons and the convergence of learning

UNIT II: (14 **Hrs**)
ML Classifiers using sckikit-learn: Choosing a classification algorithm - Training a perceptron - Modeling class probabilities via logistic regression - Maximum margin classification with support vector machines (SVM) - Solving nonlinear problems using a kernel SVM - Decision tree learning - K-nearest neighbors: lazy learning algorithm. **Data Preprocessing -** Missing data - Categorical data - Partitioning a dataset into separate training and test datasets - Bringing features onto the same scale - Selecting meaningful features - Assessing feature importance with random forests

UNIT III: **(14 Hrs)**
Compressing Data via Dimensionality Reduction: Unsupervised dimensionality reduction via principal component analysis - Supervised data compression via linear discriminant analysis - Using kernel principal component analysis for nonlinear mappings. **Learning Best Practices for Model Evaluation and Hyperparameter Tuning -** Streamlining workflows with pipelines - Using k-fold cross-validation to assess model performance - Debugging algorithms with learning and validation curves - Fine-tuning ML models via grid search - Looking at different performance evaluation metrics

UNIT IV: **(15 Hrs)**
Combining different models for ensemble learning: Learning with ensembles - Combining classifiers via majority vote - Bagging: building an ensemble of classifiers from bootstrap samples - Leveraging weak learners via adaptive boosting. **Predicting Continuous Target Variables with Regression Analysis** - Introducing Linear regression **-** Implementing an

ordinary least squares linear regression model - Fitting a robust regression model using RANSAC - Evaluating the performance of linear regression models - Using regularized methods for regression - Turning a linear regression model into a curve -polynomial regression - Dealing with nonlinear relationships using random forests

UNIT V: (15 Hrs)

Working with Unlabeled Data: Grouping objects by similarity using k-means - Organizing clusters as a hierarchical tree - Locating regions of high density via DBSCAN. **Introduction to Embedding a ML model into a Web Application -** Serialising fitted scikit-learn estimators –Develop web application using ML model - Developing a web application with Flask - Turning any classifier into a web application - Deploying the web application to a public server.

Text Books:

1. Corey Wade, Mario Corchero Jiménez, Andrew Bird, Dr. Lau Cher Han and Graham Lee,The Python Workshop, 2nd Edition, packs publishing, 2022
   Unit I:Ch. 10
2. Sebastian Raschka, Python Machine Learning, 3rd ed, packt publishing, 2019
   Unit I: Ch. 1, Ch. 2, Unit II: Ch. 3, Ch. 4, Unit III: Ch. 5, Ch. 6, Unit IV: Ch. 7, Ch. 10, Unit V: Ch. 9, Ch. 11.

Reference Books:

1. Andreas C. Mueller, Sarah Guido. Introduction to Machine Learning with Python. O'Reilly Media, Inc., 2016.
2. 2.Ethem Alpaydin, Introduction to Machine Learning, 2nd Edition, 2010
3. 3.Wes McKinney. Python for Data Analysis. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, second edition, 2018

Web References:

- https://data-flair.training/blogs/machine-learning-tutorial/
- https://www.geeksforgeeks.org/machine-learning/

MACHINE LEARNING - LAB
(Course code : 23PDSC32)

SEMESTER– III	HOURS – 6	CREDITS – 5

Objective:
To preprocess the data and build ML models using appropriate techniques and evaluate the model.

Outcomes:
At the end of the course the students will be able to:
CO1: Apply pandas, NumPy and Matplotlib to read in, process and visualize data, implement linear classification algorithms. (K3)
CO2: Compare classifiers with linear and non-linear decision boundaries, select relevant features for the model construction. (K2)
CO3: Apply data compression and best practices for model evaluation and hyper parameter tuning. (K3)
CO4: Select appropriate algorithms and ensemble. (K4)
CO5: Apply clustering algorithms on unlabeled data. (K3)
CO6: Evaluate ML models by developing web applications. (K5)

Unit I: **(14 Hrs)**
1. Programs using NumPy and pandas
2. Data Visualization
3. Ordinary least squares linear regression model
4. Evaluating the performance of linear regression models
5. Modeling class probabilities via logistic regression
6. Regularized methods for regression
7. Nonlinear relationships using random forests
8. Decision tree

Unit II: **(14 Hrs)**
1. Perceptron learning algorithm
2. Adaline
3. Multilayer Perceptron
4. Maximum margin classification with Support Vector Machines(SVM)
5. Solving nonlinear problems using a kernel SVM

Unit III: **(14 Hrs)**
1. Unsupervised dimensionality reduction via principal component analysis
2. Supervised data compression via linear discriminant analysis
3. Using k-fold cross-validation to assess model performance
4. Debugging algorithms with learning and validation curves
5. Fine-tuning ML models via grid search
6. Implementing different performance evaluation metrics

Unit IV: **(15 Hrs)**
1. Ensemble Learning

Unit V: **(15 Hrs)**
2. Grouping objects by similarity using k-means
3. Organizing clusters as a hierarchical tree
4. Locating regions of high density via DBSCAN
5. Embedding a ML model into a Web Application

Text Books:

1. Corey Wade, Mario Corchero Jiménez, Andrew Bird, Dr. Lau Cher Han and Graham Lee,The Python Workshop, 2nd Edition, packs publishing, 2022.
2. Sebastian Raschka, Python Machine Learning, 3rd Edition, packt publishing, 2019.

Reference Book:

1. Andreas C. Mueller, Sarah Guido. Introduction to Machine Learning with Python. O'Reilly Media, Inc., 2016.
2. 2.Ethem Alpaydin, Introduction to Machine Learning, 2nd Edition, 2010.
3. 3.Wes McKinney. Python for Data Analysis. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, 2ndedition, 2018

Web Resources:

- https://machinelearningmastery.com/machine-learning-in-python-step-by-step/
- https://www.tutorialspoint.com/machine_learning_with_python/index.htm
- https://pythonprogramming.net/machine-learning-tutorial-python-introduction/

DATA BASES FOR DATA SCIENCE
(Course code : 23PDSC33)

SEMESTER– III	HOURS – 6	CREDITS – 5

Objectives:
To provide fundamentals of database design, modeling systems, data storage, world of data warehousing and NoSQL.

Outcomes:
At the end of the course the students will be able to:
CO1: Discuss the importance of relational data modeling and conceptual Modelling. (K2)
CO2: Experiment with various database and compose effective queries. (K3)
CO3: Analyze the process of OLAP system construction. (K4)
CO4: Evaluate the use of NOSQL and its approach to the database. (K5)
CO5: Develop applications using Relational and NoSQL databases. (K6)
CO6: Create a database using NoSQL. (K6)

UNIT I: (14 Hrs)
Fundamental Concepts of Database Management: Applications of Database Technology - Key Definitions - File versus Database Approach to Data Management - Elements of a Database System - Advantages of Database Systems and Database Management - Architecture and Categorization of DBMSs. **Conceptual Data Modelling using the ER Model and UML Class Diagram:** Phases of Database Design - The Entity Relationship Model - UML Class Diagram

UNIT II: (14 Hrs)
Types of Database Systems: Legacy Databases - Relational Databases: The Relational Model – Normalization. **Relational Databases:** Structured Query Language - SQL Data Definition Language - SQL Data Manipulation Language.
Lab: SQL DDL and DML**.**

UNIT III: (14 Hrs)
Data Warehousing and Business Intelligence: Operational versus Tactical/Strategic Decision-Making - Data Warehouse Definition - Data Warehouse Schemas - The Extraction, Transformation, and Loading (ETL) Process - Data Marts - Virtual Data Warehouses and Virtual Data Marts - Operational Data Store - Data Warehouses vs Data Lakes - Business Intelligence. **Introduction of NO SQL:** Overview and History of NoSQL Databases. Definition of the Four Types of NoSQL Database, The Value of Relational Databases, Getting at Persistent Data, Concurrency, Integration, Impedance Mismatch, Application and Integration Databases, Attack of the Clusters, The Emergence of NoSQL, Key Points Comparison of relational databases to new NoSQL stores, Mongo DB, Cassandra, HBASE, Neo4j use and deployment, Application, RDBMS approach, Challenges NoSQL approach, Key-Value and Document Data Models, Column-Family Stores, Aggregate Oriented Databases. sharding, MapReduce on databases. Distribution Models, Single Server, Sharding, Master-Slave Replication, Peer-to-Peer replication, Combining Sharding and Replication.

UNIT IV: (15 Hrs)
Key Value Data Stores: NoSQL Key/Value databases using MongoDB, Document Databases, Document oriented Database Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Event Logging, Content Management Systems, Blogging Platforms, Web Analytics or Real-Time Analytics, E-Commerce Applications, Complex Transactions Spanning Different Operations, Queries against Varying Aggregate Structure.

Lab: Key-value databases, Replica of existing database, Backup of existing database, Restore database from the backup
Demonstration: Connecting python with mongodb and inserting, retrieving, updating, and deleting.

UNIT V: (15 Hrs)

Document Oriented Database: Column- oriented NoSQL databases using Apache HBASE, Column-oriented NoSQL databases using Apache Cassandra, Architecture of HBASE, Column-Family Data Store Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Event Logging, Content Management Systems, Blogging Platforms, Counters, Expiring Usage. **Data Modeling with Graph:** Comparison of Relational and Graph Modeling, Property Graph Model Graph Analytics: Link analysis algorithm- Web as a graph, Page Rank Markov chain, page rank computation, Topic specific page rank Page Ranking Computation techniques iterative processing, Random walk distribution Querying Graphs.
Lab: Implement with column-family stores (cassandra), Graph databases (neo4j), Aggregate function, Push and addtoset expression, First and last expression.

Text Books:

1. Lemahieu, W., Broucke, S.Vanden, Baesens, B. (2018) Principles of database management:The Practical Guide to Storing, Managing and Analyzing big and small data. Cambridge, UnitedKingdom: Cambridge University Press.
2. Sadalage, P., Fowler,NoSQL Distilled: A Brief Guide to the Emerging World of PolyglotPersistence, Wiley Publications,1st Edition,2022

Reference Books:

1. SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis Renee M. P. Teate
2. SQL for Data Science: Data cleaning, wrangling and analytics with relational databases, Antonio Badia
3. Guy Harrison, Next Generation Database: NoSQL and big data, Apress

Web References:

- https://www.geeksforgeeks.org/introduction-to-nosql/

NATURAL LANGUAGE PROCESSING
(Course code : 23PDSC34)

SEMESTER– III	HOURS – 6	CREDITS – 4

Objective:

To explore the concepts and fundamentals of Natural Language Programming.

Outcome:

At the end of the course the students will be able to

CO1: Summarize the fundamentals of Natural Language Processing. (K2)
CO2: Apply the NLP techniques for word and syntactic analysis. (K3)
CO3: Analyze the text using text editors. (K4)
CO4: Evaluate the tools and methods for understanding semantics of sentences And pragmatics.(K5)
CO5: Develop an innovative application using NLP components. (K6)
CO6: Evaluate an application using NLP components. (K5)

UNIT-I: (7 Hrs)

INTRODUCTION TO NLP: Knowledge in Speech and Language Processing – Ambiguity - Models and Algorithms- Language, Thought, and Understanding - The State of the Art - History - Applications – Basic NLP.

UNIT-II: (7 Hrs)

WORD ANALYSIS: Regular Expressions - Words & Transducers- Survey of English Morphology- Finite-State Morphological Parsing - Word and Sentence Tokenization- N-grams- Counting Words in Corpora- Simple (Unsmoothed) N-grams- Training and Test Sets- Part-of-Speech Tagging- English Word Classes- Tagsets for English- Part-of-Speech Tagging- Rule-Based Part-of-Speech Tagging- Evaluation and Error Analysis.

UNIT-III: (7 Hrs)

SYNTACTIC ANALYSIS: Formal Grammars of English- Constituency- Context-Free Grammars- Grammar Rules for English - Treebanks - Finite-State and Context-Free Grammars - Dependency Grammars - Parsing with Context-Free Grammars - Parsing as Search- Dynamic Programming Parsing Methods- Statistical Parsing- Probabilistic Context-Free Grammars- Probabilistic CKY Parsing of PCFGs- Learning PCFG Rule Probabilities.

UNIT-IV: (7 Hrs)

SEMANTICS AND PRAGMATICS: Computational Desiderata for Representations- First-Order Logic, Computational Semantics –Syntax Driven Semantic analysis, Semantic attachments Semantic Attachments for a Fragment of English, Lexical Semantics- Word Senses, Relations between Senses, WordNet: A Database of Lexical Relations- Event Participants: Semantic Roles and Selectional Restriction.Word/Sentence/Text embeddings—Topic Models —LSA— LDA

UNIT-V: (8 Hrs)

APPLICATIONS And Transformers: Applications - Information Extraction, Question Answering and Summarization - Dialogue and Conversational Agents (Addition). Text Classification - Transformers and Pretrained Language Models

Text Books:

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.Chapter 1,2,3,4,5,12,13,14,17,18,19,22.23.24
2. Patel, Ankur A., and Ajay Uppili Arasanipalai. Applied Natural Language Processing in the Enterprise. " O'Reilly Media, Inc.", 2021. Chapter1.

Reference Books:

1. Breck Baldwin, Language Processing with Java and LingPipe Cookbook, Atlantic Publisher, 2015.
2. Richard M Reese, Natural Language Processing with Java‖, O_Reilly Media, 2015.
3. Nitin Indurkhya and Fred J. Damerau, Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.
4. Tanveer Siddiqui, U.S. Tiwary, Natural Language Processing and Information Retrieval‖, Oxford University Press, 2008.

Web References:

- https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/what-is-natural-language-processing-nlp
- https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1
- https://www.oracle.com/in/artificial-intelligence/what-is-natural-language-processing/

PREDICTIVE DATA ANALYTICS
(Course code : 23PDSE31)

SEMESTER– III	HOURS – 4	CREDITS – 3

Objective:

To enable to develop a deep understanding of predictive analytics principles, statistical methodologies, and apply advanced techniques, and formulate innovative solutions aligned with current trends.

Outcomes:

At the end of the course the students will be able to

CO1: Describe the foundational concepts of data analytics, explain the challenges of conventional systems, and illustrate the methodologies used in intelligent data analysis. (K1)

CO2: Apply quantitative analysis techniques including probability distributions, inferential statistics, and hypothesis testing, using tools such as Rand Python. (K2)

CO3: Utilize SQL and data importing techniques to perform data mining and analysis of variance, while employing statistical process control and reporting methods. (K3)

CO4: Examine and interpret various statistical models, including explanatory and predictive models, and demonstrate competence in implementing machine learning methodologies.(K4)

CO5: Critically analyze components, and applications, and effectively work with data to improve decision-making and performance. (K5)

CO6: Formulate innovative solutions related to prediction for the challenges real time, leveraging advanced knowledge of statistical methodologies and tools.(K6)

UNIT I: (14 Hrs)

Linear Methods for Regression and Classification: Overview of supervised learning– Linear regression models and least squares, Multiple regression, Multiple outputs, Subset selection – Ridgeregression - Lasso regression - Linear Discriminant Analysis - Logistic regression - Perceptron learning algorithm.

UNIT II: (14 Hrs)

Model Assessment and Selection: Bias – Varianceand model complexity – Bias – variancetrade off - Optimism of training error rate - Estimate of In-sample prediction error - Effective number of parameters – Bayesian approach and BIC - Cross- validation – Bootstrap methods – conditional or expected test error

UNIT III: (14 Hrs)

Additive Models, Trees, and Boosting: Generalized additive models - Regression and classification – trees – Boosting methods – exponential loss and Ada Boost - Numerical Optimization via gradient Boosting - Examples - California housing – NewZealand fish - Demographic data.

UNIT IV: (15 Hrs)

Neural Networks: Fitting neural networks - Back propagation - Issues in training NN - SVM for classification - Reproducing Kernels - SVM for regression – Knearest Neighbour classifiers - Image Scene Classification.

UNIT V: (15 Hrs)

Unsupervised Learning and Random forests: Association rules - Cluster analysis– Principal Components - Random forests and analysis. **Model Evaluation and Deployment:** Model Validation - Rule Induction - Using CHAID - Automating Models - Comparing and

Combining Models - Evaluation Charts – Meta Level Modeling, Deploying Model, Assessing Model Performance, Updating a Model.

Text Books:

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag,2009.
2. E.Alpaydin, Introduction to Machine Learning, Prentice Hall of India, 2010.
3. Eric Siegel, Predictive analytics, Wiley, 2016.

Reference Books:

- C.M.Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
- G. James, D. Witten, T. Hastie, R. Tibshirani, An introduction to statistical learning with applications in R, Springer, 2013.
- Jensen, C., Data Science for Business: Data Analytics Guide with Strategies and Techniques, 2017.

SOCIAL NETWORK ANALYSIS
(Course code : 23PDSE31)

SEMESTER– III	HOURS – 4	CREDITS – 3

Objective:
To introduce the concepts and fundamentals of social network components and analysis.

Outcomes:
At the end of the course the students will be able to
CO1 : Differentiate the fundamentals of social web and elements of social network analysis. (K2)
CO2 : Apply and visualize the knowledge representation in social network. (K3)
CO3 : Analyze the various methods in social network analysis. (K4)
CO4 : Evaluate the tools and methods for analyzing the social network data. (K4)
CO5 : Investigate the recent potential applications and development of social network with real time case studies. (K4)
CO6 : Evaluate the data using semantic tools. (K4)

UNIT-I: (7 Hrs)
Introduction to Semantic Web and Social Networks: Introduction to Semantic Web: Limitations of current Web - Development of Semantic Web - Emergence of the Social Web - Social Network analysis: Development of Social Network Analysis - Key concepts and measures in network analysis - Electronic sources for network analysis: Electronic discussion networks, Blogs and online communities - Web-based networks - Applications of Social Network Analysis- Brief history of social network analysis.

UNIT-II: (7 Hrs)
Modelling, Aggregating and Knowledge Representation: Knowledge Representation on the semantic web- Ontology and their role in the Semantic Web - Ontology languages for the Semantic Web- Modelling and aggregating social network data: State-of-the-art in network data representation - Ontological representation of social individuals - Ontological representation of social relationships - Aggregating and reasoning with social network data - Advanced representations.

UNIT-III: (7 Hrs)
Data Collection: Boundary specification – Data collection process- Information bias and issue of reliability – Archival data – Understanding SNA data – Managing SNA data.

UNIT-IV: (8 Hrs)
Methods in Social Network Analysis: Descriptive methods – Graph – Density- Centrality – cliques – MDS- structural equivalence – Two mode networks – Inferential methods – QAP-ERGM.

UNIT-V: (7 Hrs)
Case Studies: Case studies – Evaluation of web-based social network extraction – semantic – based social network analysis in the sciences – emergent semantics.

Text Book:
1. Peter Mika, "Social Networks and the Semantic Web", Springer 2007.
2. Yang, Song, Franziska B. Keller, and Lu Zheng. Social network analysis: Methods and examples. Sage Publications, 2016.

Unit I: *Book 1- Chapter 1,2,3 & Book 2: Chapter 1,* ***Unit II:*** *Book 1 - Chapter 4,5,6*
Unit III: *Book2 - Chapter 2,* ***Unit IV:*** *Book 2- Chapter 3, 4,* ***Unit V:*** *Book 1- Chapter 7,8,9*

Reference Book:

1. GuandongXu, Yanchun Zhang and Lin Li, Web Mining and Social Networking – Techniques and applications, First Edition, Springer, 2011.
2. Dion Goh and Schubert Foo, Social information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively‖, IGI Global Snippet, 2008.

Web References:

- https://bookdown.org/chen/snaEd/ch4.html
- https://www.sciencedirect.com/topics/social-sciences/social-network-analysis
- https://www.publichealth.columbia.edu/research/population-health-methods/social-network-analysis
- https://www.ibm.com/docs/en/spss-modeler/18.0.0?topic=analysis-about-social-network

REINFORCEMENT LEARNING
(Course code : 23PDSE31)

SEMESTER– III	HOURS – 4	CREDITS – 3

Objective:

Acquire a comprehensive understanding of Reinforcement Learning concepts and methodologies, progressing from foundational principles and basic problem-solving methods to advanced techniques and real-world applications.

Outcomes:

At the end of the course the students will be able to

CO1: Associate the fundamentals of Reinforcement Learning tasks and the core principals including policies, value, and functions. (K1)

CO2: Apply the tabular and approximation methods to solve classical control problems. (K2)

CO3: Analyze policy gradient methods to solve more complex cases. (K3)

CO4: Evaluate the tools and methods used for prediction and control. (K4)

CO5: Investigate the current advanced techniques and applications in Reinforcement Learning. (K5)

CO6: Evaluate the data using semantic tools.(K6)

UNIT-I **(12 Hrs)**

INTRODUCTION AND BASICS OF REINFORCEMENT LEARNING: The Reinforcement Learning Problem - Reinforcement Learning- Examples- Elements of Reinforcement Learning- Limitations and Scope -An extended example – History of Reinforcement Learning - Applications - Ethics in RL- Applying RL for real-world problems- Meta-learning- Multi-Agent Reinforcement Learning.

UNIT-II: **(12 Hrs)**

TABULAR METHODS: Finite Markov Decision Processes - Dynamic Programming - Monte Carlo Methods.

UNIT-III: **(12 Hrs)**

Q-NETWORKS AND LEARNING: Temporal difference learning – n-step Bootstrapping- Planning and learning with tabular methods, Deep Q-networks- DQN, DDQN, Dueling DQN, Prioritised Experience Replay.

UNIT-IV: **(12 Hrs)**

APPROXIMATE SOLUTION METHODS: On-policy prediction with approximation – on-policy control with approximation – policy gradient methods

UNIT-V: CASE STUDIES **(12 Hrs)**

PSYCHOLOGY AND NEUROSCIENCE: Prediction and control - Classical conditioning – neuroscience – basics- reward and prediction -case studies.

Text Book:

1. Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.

Unit I: Ch. 1, **Unit II:**Ch. 3,4,5, **Unit III:**Ch. 6,7,8, **Unit IV:**Ch. 9,10,13, **Unit V:**Ch. 14,15,16

Reference Books:

1. Szepesvári, Csaba. "Algorithms for reinforcement learning." Synthesis lectures on artificial intelligence and machine learning 4.1 (2010).

2. Winder, Phil. Reinforcement learning. O'Reilly Media, 2020.
3. Bilgin, Enes. Mastering Reinforcement Learning with Python: Build next-generation, self-learning models using reinforcement learning techniques and best practices. Packt Publishing Ltd, 2020.

Web References:

- https://developer.ibm.com/learningpaths/get-started-automated-ai-for-decision-making-api/what-is-automated-ai-for-decision-making/
- https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292
- https://www.analyticsvidhya.com/blog/2021/02/introduction-to-reinforcement-learning-for-beginners/

BUSINESS AND HEALTH CARE ANALYTICS
(Course code : 23PDSE31)

SEMESTER– III	HOURS – 4	CREDITS – 3

Objective:

To enable to develop a deep understanding of data analytics principles, statistical methodologies, and apply advanced techniques, and formulate innovative solutions aligned with current trends in business and healthcare analytics.

Outcomes:

At the end of the course the students will be able to

CO1: Describe the foundational concepts of data analytics, explain the challenges of conventional systems, and illustrate the methodologies used in intelligent data analysis. (K1)

CO2: Apply quantitative analysis techniques including probability distributions, inferential statistics, and hypothesis testing, using tools such as Excel, R, STATA, PSPP, and EVIEWS. (K2)

CO3: Utilize SQL and data importing techniques to perform data mining and analysis of variance, while employing statistical process control and reporting methods. (K3)

CO4: Examine and interpret various statistical models, including explanatory and predictive models, and demonstrate competence in implementing machine learning methodologies. (K4)

CO5: Critically analyze business/ healthcare analytics components, and applications, and effectively work with data to improve decision-making and performance. (K5)

CO6: Formulate innovative solutions for challenges in the field of healthcare and business, leveraging advanced knowledge of statistical methodologies and tools. (K6)

UNIT I **(14 Hrs)**

Introduction to Data Analytics Platform - Visualizing Data - Describing and Summarizing Data - Challenges of Conventional Systems - Intelligent Data Analysis - Analytic Methodologies or Techniques Used in Logical Analysis.

UNIT II **(14 Hrs)**

Quantitative Analysis - Sampling Methods and Estimation – Probability Distributions - Descriptive Statistics - Inferential Statistics - Hypothesis Testing, Explanatory and Predictive Models, and Fact-Based Management to Drive Decisions and Actions - Tools - Analysis vs Reporting.

UNIT III **(14 Hrs)**

One-Sample Tests - Two Independent Samples Tests - K Related Samples Tests - Measures of Correlation and Association - Multivariate Nonparametric Test for Interdependence - Probability and Decision Making Under Uncertainty - Normal, Binomial, Poisson, and Exponential Distributions.

UNIT IV **(15 Hrs)**

Data Mining - Importing Data into Excel - SQL - Analysis of Variance and Experimental Design - Statistical Process Control - Statistical Reporting - Foundations, Methods, Interpretations in Excel – R – STATA – PSPP – EVIEWS – Machine Learning.

UNIT V **(15 Hrs)**

Fundamental of Health care Analytics-Improve Decision Making-Analytics, Quality and Performance-Applications-Components-Working with data-Raw material and Analyzing data.

List of Activities

1. Using Excel/R, load a sample dataset and demonstrate how to perform basic data manipulation tasks like sorting, filtering, and calculating summary statistics.
2. Utilizing the chosen platform, analyze a healthcare dataset to calculate mean, median, standard deviation, and create a frequency distribution for a specific variable.
3. Incorporate the selected platform to conduct a hypothesis test comparing the means of two health care-related groups, calculating the p-value and confidence interval.
4. Using the analytics tool, simulate a binomial distribution to model the probability of successful surgeries in a healthcare context, and visualize the results.
5. Write an SQL query to extract patient data from a healthcare database, filtering for specific conditions such as age range and medical history.
6. Analyze a healthcare dataset with multiple treatment groups, using ANOVA via the analytics tool to determine if there are significant differences in treatment outcomes.
7. Apply a simple linear regression model using R/Python to predict patient hospitalization time based on variables like age and medical history.
8. Explore a healthcare dataset and identify trends in-patient admissions based on age, gender, and diagnosis, presenting your findings visually using the analytics platform.
9. Using the chosen tool, analyze a dataset containing blood pressure measurements over time and apply control limits to identify potential anomalies.
10. Choose a healthcare dataset, perform in-depth analysis using your preferred analytics platform, and create a comprehensive report with visualizations and insights on patient demographics, treatments, and outcomes.

Text Books:

1. Bowerman, B. (2016). Business Statistics in Practice: Using Data, Modeling, and Analytics. McGraw-Hill Higher Education.
2. Christian Albright, Wayne L. Winston (2015). Business Analytics: Data Analysis and DecisionMaking, 5thEdition, CENGAGE
3. Prasad R N and Seema Acharya (2016). Fundamentals of Business Analytics, 2ed, Wiley
4. Trevor Strome,Health care Analytics for quality and Performance improvement,Wiley,2013.

Reference Books:

1. Jensen, C. (2017). Data Science for Business: Data Analytics Guide with Strategies and Techniques
2. Gert H. N. Laursen, JesperThorlund (2018). Business Analytics for Managers, 2ed: Taking Business Intelligence Beyond Reporting, Wiley.
3. Reddy Chandan K, and Charu C.Aggarwal,eds,Healthcare Data Analytics.

Web Resources:

- https://www.simplilearn.com/resources/data-science-business-analytics

PROFESSIONAL COMMUNICATION SKILL
(Course code : 23PDSS31)

SEMESTER– III	HOURS – 2	CREDITS – 2

Objectives:

The course aims to:

1. Enhance the Employability and Career Skills of students
2. Orient the students towards grooming as a professional
3. Make them Employability Graduates
4. Develop their confidence and help them attend interviews successfully.

Outcomes:

At the end of the course learners will be ability to:

CO1: Illustrate Communication Strategies (K4)
CO2: Make effective presentations (K6)
CO3: Participate confidently in Group Discussions. (K6)
CO4: Attend job interviews and be successful in them. (K6)
CO5: Develop adequate Soft Skills required for the workplace (K6)
CO6: Develop skill to manage stress and long-term career plan (K6)

UNIT I **(10 Hrs)**

Introduction to Soft Skills- Hard skills & soft skills - employability and career Skills—Grooming as a professional with values—Time Management—General awareness of Current Affairs.

UNIT II **(10 Hrs)**

Self-Introduction-organizing the material - Introducing oneself to the audience – introducing the topic – answering questions – individual presentation practice— presenting the visuals effectively – 5-minute presentations – writing: email – web content – Preparation of CV.

UNIT III **(9 Hrs)**

Introduction to Group Discussion— Participating in group discussions – understanding group dynamics - brainstorming the topic -– questioning and clarifying –GD strategies- activities to improve GD skills.

UNIT IV **(9 Hrs)**

Interview etiquette – dress code – body language – attending job interviews– telephone/skype interview -one to one interview &panel interview – FAQs related to job interviews.

UNIT V **(10 Hrs)**

Recognizing differences between groups and teams- managing time-managing stress-networking professionally- respecting social protocols-understanding career management-developing a long- term career plan-making career changes.

Reference Books:

1. Butterfield, Jeff Soft Skills for Everyone. Cengage Learning: New Delhi, 2015
2. Interact English Lab Manual for Undergraduate Students, Orient Black Swan: Hyderabad, 2016.
3. E. Suresh Kumar et al. Communication for Professional Success. Orient Blackswan: Hyderabad, 2015.

Web References:

- https://novoresume.com/career-blog/communication-skills
- https://in.indeed.com/career-advice/resumes-cover-letters/communication-skills
- https://study.com/learn/lesson/professional-communication-skills.html

INTERNSHIP
(Course code : 23PDSI35)

SEMESTER– III	INTERNSHIP	CREDITS –2

- All PG students will undergo internship during the summer holidays of the First year after completing II semester.
- Two credits will be given for internship.
- Minimum Days: 30
- Minimum working time per day: 3 Hrs. & Maximum working Time: 5 Hrs.
- The places of internship can be government offices, Panchayats, MP, MLA offices, private institutions, companies, production units etc.
- The HoD of the departments will give a letter of introduction to each student.
- The students will identify the company / institution for internship.
- The students will be divided equally based on the number of professors available in the departments. Each professor will serve as a guide to the assigned students.
- The students will finalize the institutions / companies for the internship in consultation with the guides.
- The students shall maintain a work diary which will be countersigned by the managers / authorities of the company in which the students do the internship on daily basis.
- The work diary, Work completion certificate obtained from the company and a comprehensive report on the learning outcomes will be submitted to the guides at the end of the internship.
- Viva will be conducted based on the experience of the internship in the month of August. The guide will be the internal examiner and another faculty from the same department will serve as the external examiner.

DEEP LEARNING
(Course code: 23PDSC41)

SEMESTER– IV	HOURS –6	CREDITS – 5

Objective:

To develop the ability to construct efficient neural networks and apply deep learning techniques for real-time applications, combining theoretical understanding with practical implementation.

Outcomes:

At the end of the course Learners will be ability to

CO1: Illustrate the major technology trends in neural networks and deep learning. (K4)
CO2: Analyse neural networks and fully connected deep neural networks (K4)
CO3: Apply neural networks and fully connected deep neural networks. (K3)
CO4: Evaluate efficient (vectorized) neural networks and deep learning for real time application. (K6)
CO5: Build efficient (vectorized) neural networks and deep learning for real time application. (K5)
CO6: Create DL models using CNN. (K5)

UNIT-I: (10 Hrs)

Introduction to Artificial Neural Networks: Neural Networks-Application Scope of Neural Networks- Fundamental Concept of ANN: The Artificial Neural Network-Biological Neural Network-Comparison between Biological Neuron and Artificial Neuron-Evolution of Neural Network. Basic models of ANN-Learning Methods-Activation Functions-Importance Terminologies of ANN.

UNIT-II: (10 Hrs)

Supervised Learning Network - Shallow neural networks- Perceptron Networks-Theory-Perceptron Learning Rule Architecture-Flowchart for training Process-Perceptron Training Algorithm for Single and Multiple Output Classes. Back Propagation Network- Theory-Architecture-Flowchart for training process -Training Algorithm-Learning Factors for Back-Propagation Network. Radial Basis Function Network RBFN: Theory, Architecture, Flowchart and Algorithm.

UNIT-III: (10 Hrs)

Convolutional Neural Network: Introduction - Components of CNN Architecture - Rectified Linear Unit (ReLU) Layer - Exponential Linear Unit (ELU, or SELU) - Unique Properties of CNN -Architectures of CNN -Applications of CNN.

UNIT-IV: (10 Hrs)

Recurrent Neural Network: Introduction- The Architecture of Recurrent Neural Network-The Challenges of Training Recurrent Networks- Echo-State Networks- Long Short-Term Memory (LSTM) - Applications of RNN.

UNIT-V: (10 Hrs)

Auto Encoder and Restricted Boltzmann Machine: Introduction - Features of Auto encoder Types of Autoencoder Restricted Boltzmann Machine- Boltzmann Machine - RBM Architecture - Types of RBM.

Text Books:

1. S. N. Sivanandam, S. N. Deepa, Principles of Soft Computing, Wiley-India, 3rd Edition, 2018.

2. S Lovelyn Rose, L Ashok Kumar,D Karthika Renuka, Deep Learning using Python, Wiley-India, 1st Edition, 2019.

Reference Books:

1. Charu C. Aggarwal, Neural Networks and Deep Learning, Springer, September 2018.
2. Francois Chollet, Deep Learning with Python, Manning Publications; 1st edition,2017
3. John D. Kelleher, Deep Learning (MIT Press Essential Knowledge series), The MIT Press, 2019.

Web References:

- https://onlinecourses.nptel.ac.in/noc22_cs22/preview
- https://arxiv.org/abs/1506.06579
- https://arxiv.org/abs/1605.06211
- https://cs230.stanford.edu/lecture/

BIG DATA ANALYTICS
(Course code: 23PDSC42)

SEMESTER– IV	HOURS –6	CREDITS – 5

Objective:
To introduce the concepts of big data analytics and developing a real time application.

Outcomes:
At the end of the course Learners will be ability to
CO1: Restate the basic concepts of big data analytics and technologies. (K2)
CO2: Apply the concept of HDFS, Map reduce for storing and processing of Big Data. (K3)
CO3: Analyze and perform different operations on data using Pig, Hive, and Hbase. (K4)
CO4: Evaluate the tools and methods for analyzing Big Data analytics model. (K5)
CO5: Develop real time big data analytics applications. (K6)
CO6: Create an application using recent tools. (K6)

UNIT-I: (10 Hrs)
INTRODUCTION TO BIG DATA ANALYTICS: Classification of Digital Data, Structured and Unstructured Data - Introduction to Big Data: Characteristics – Evolution – Definition - Challenges with Big Data - Other Characteristics of Data - Why Big Data - Traditional Business Intelligence versus Big Data - Data Warehouse and Hadoop Environment Big Data Analytics: Classification of Analytics – Challenges - Big Data Analytics important - Data Science - Data Scientist - Terminologies used in Big Data Environments.

UNIT-II: (10 Hrs)
BIG DATA TECHNOLOGY LANDSCAPE: NoSQL, Comparison of SQL and NoSQL, Hadoop -RDBMS Versus Hadoop - Distributed Computing Challenges – Hadoop Overview - Hadoop Distributed File System - Processing Data with Hadoop - Managing Resources and Applications with Hadoop YARN - Interacting with Hadoop Ecosystem.

UNIT-III: (10 Hrs)
HADOOP AND HDFS: Introduction to Hadoop – RDBMS vs Hadoop- distributed computing challenges - A Brief History of Hadoop- The Hadoop Distributed Filesystem- Processing Data with Hadoop - Anatomy of a MapReduce Works - Anatomy of a MapReduce Job Run- Job Scheduling- Shuffle and Sort- Task Execution.

UNIT-IV: (9 Hrs)
Hive: Introduction – Architecture - Data Types - File Formats - Hive Query Language Statements – Partitions – Bucketing – Views - Sub- Query – Joins – Aggregations - Group by and Having - RCFile Implementation - Hive User Defined Function - Serialization and Deserialization. Pig: Introduction - Anatomy – Features – Philosophy - Use Case for Pig - Pig Latin Overview - Pig Primitive Data Types - Running Pig - Execution Modes of Pig - HDFS Commands - Relational Operators - Eval Function - Complex Data Types - Piggy Bank - User-Defined Functions - Parameter Substitution - Diagnostic Operator - Word Count Example using Pig - Pig at Yahoo! - Pig Versus Hive Hbase - HBasics, PySparkConcepts.

UNIT-V: (9 Hrs)
Case Studies: Hadoop Usage at Last.fm - Hadoop and Hive at Facebook- Nutch Search Engine- Log Processing at Rackspace – Cascading –TeraByte Sort on Apache Hadoop - Using Pig and Wukong to Explore Billion-edge Network Graphs - Recent Trends in Big Data Analytics.

Text Book:

1. Seema Acharya, Subhashini Chellappan, Big Data and Analytics, Wiley, First Edition, 2015.
2. Tom White, Hadoop: The Definitive Guide, O'Reilly Media Inc., 2015.

Unit I: *Book 1 – Ch. 1,2,3,* ***Unit II:*** *Book 1: Ch. 4, 5,* ***Unit III:*** *Book 2 – Ch.1, 3,6*
Unit IV: *Book 1 – Ch. 9, 10, Book 2 – Ch. 11, 12,13.* ***Unit V:*** *Book 2 - Ch. 16*

Reference Books:

1. Lublinsky, Boris, Kevin T. Smith, and Alexey Yakubovich. Professional hadoop solutions. John Wiley & Sons, 2013.
2. RadhaShankarmani, M Vijayalakshmi, Big Data Analytics, 2nd Edition, 2017, Wiley
3. Henry H. Liu, Hadoop Essentials: A Quantitative Approach, 1st Edition, 2012, PerfMath Publishers

Web References:

- https://www.ibm.com/analytics/big-data-analytics
- https://www.simplilearn.com/what-is-big-data-analytics-article
- https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-big-data-analytics

BIG DATA ANALYTICS LAB
(Course code: 23PDSC42)

SEMESTER– IV	HOURS – 6	CREDITS – 5

Objective:
To introduce the concepts of big data analytics and developing a real time application.

Outcomes:
At the end of the course Learners will be ability to:
CO1: Configure Hadoop and perform File Management (K6)
CO2: Apply Map Reduce program to real time issues. (K3)
CO3: Critically analyze huge data set using Hadoop distributed file systems and Map Reduce (K4)
CO4: Experimenting different data processing tools like Pig, Hive. (K1)
CO5: Develop real time big data analytics applications (K6)
CO6: Create and evaluate a application using tools. (K6)

UNIT-I **(10 Hrs)**
1. Install Apache Hadoop
2. Perform setting up and Installing Hadoop in its three operating modes: Standalone, Pseudo Distributed and Fully Distributed
3. To use Web Based Tools to Manage Hadoop Set-up
4. Implement the following file management tasks in Hadoop: adding files and directories, retrieving files and deleting Files

UNIT-II **(10 Hrs)**
5. Develop a Map Reduce program to calculate the frequency of a given word in a given file.
6. Develop a Map Reduce program to find the maximum temperature in each year.
7. Develop a Map Reduce program to find the grades of students.
8. Develop a Map Reduce program to implement Matrix Multiplication.
9. Develop a Map Reduce to find the maximum electrical consumption in each year given electrical consumption for each month in each year.

UNIT-III **(10 Hrs)**
10. Develop a Map Reduce to analyze weather data set and print whether the day is shinny or cool day. (National Climatic Data Centre (NCDC) Data set)
11. Develop a Map Reduce program to find the number of products sold in each country by considering sales data containing fields like Tranction _Date, Product Price, Payment Type Name, City\State, Country, Account Created, Last Login, Latitude Login
12. Data sets from different sources as Input
13. Develop a Map Reduce program to find the tags associated with each movie by analyzing movie lens data. (https://www.kaggle.com/datasets/grouplens/movielens-20m-dataset)
14. Sorting the data using Map Reduce
15. Count the number of missing and invalid values through joining two large given datasets.

UNIT-IV **(9 Hrs)**
16. Install and Run Pig then write Pig Latin scripts to sort, group, join, project and filter the data.
17. Install and Run Hive then use Hive to Create, alter and drop databases, tables, views, functions and Indexes.

18. Develop a program to calculate the maximum recorded temperature by year wise for the weather dataset in Pig Latin
19. Develop a program to calculate the maximum recorded temperature by year wise for the weather dataset in Pig Latin
20. Write queries to sort and aggregate the data in a table using HiveQL
21. Develop a Map Reduce to find the maximum electrical consumption in each year given electrical consumption for each month in each year.
22. Write a program to implement combining and partitioning in hadoop to implement a custom partitioner and Combiner

UNIT-V **(9 Hrs)**

23. Analyze the sentiment for product reviews, this work proposes a Map Reduce technique provided by Apache Hadoop
24. Trend Analysis based on Access Pattern over Web Logs using Hadoop.
25. Implementation of decision tree algorithms using Map Reduce.
26. Implementation of K-means Clustering using Map Reduce.
27. Generation of Frequent Item set using Map Reduce.

PROJECT WITH VIVA VOCE
(Course code : 23PDSC43)

SEMESTER– IV	HOURS –10	CREDITS – 7

Structure of the Project Report

1. Cover Page
2. Certificate
3. Declaration
4. Acknowledgement
5. Chapter-I Introduction
6. Chapter-II
7. Chapter-III
8. Chapter-IV
9. Chapter-V Conclusion and Scope for further research

Assessment

Internal : 100 Marks

External : 100 Marks

ARTIFICIAL INTELLIGENCE
(Course code : 23PDSE41)

SEMESTER– IV	HOURS – 4	CREDITS – 3

Objective:

To explore the approaches and principles of Artificial Intelligence (AI) algorithms, and apply them in Data Science.

Outcomes:

At the end of the course Learners will be ability to:

CO1: Contrast the problems that are amenable to solution by AI methods. (K2)
CO2: Analyse and apply appropriate AI methods to solve a given problem. (K3)
CO3: Analyse and formalize a given problem in the language/framework of different AI and learning methods. (K4)
CO4: Evaluate the AI methodologies and DL networks. (K5)
CO5: Develop AI framework to tackle projects in our increasingly complex world. (K6)
CO6: Create and evaluate the algorithm using AL methods. (K6)

UNIT-I: (10 Hrs)

Artificial Intelligence: The AI Problems - The Underlying Assumptions – What is an AI Technique – The Level of the Model – Criteria for Success. Problems, Problem Spaces & Search: Defining the problem as a State Space Search – Production systems – Problem Characteristics - Production Systems Characteristics – Issues in the Design of Search Programs. Heuristic Search Techniques: Generate and Test – Hill Climbing – Best First Search – Problem Reduction - Constraint Satisfaction – Means ends Analysis.

UNIT-II: (10 Hrs)

Knowledge Representation Issues: Representations and Mappings – Approaches to KR – Issues in KR – The Frame Problem. Using Predicate Logic - Representing Simple Facts in Logic - Representing Instances and ISA Relationships – Computable Functions and Predicates – Resolutions – Natural Deductions.

Representing Knowledge using Rules - Procedural versus Declarative Knowledge – Logic Programming – Forward Versus Backward Reasoning – Matching – Control Knowledge. Statistical Reasoning: Probability and Bayes Theorem - Certainty Factors and Rule based Systems – Bayesian Networks – Dempsters Shafer Theory - Fuzzy Logic.

UNIT-III: (10 Hrs)

Learning: What is Learning - Rote Learning – Learning by Taking Advice – Learning by Problem Solving – Learning from Examples: Induction – Explanation based Learning – Discovery – Analogy – Formal Learning Theory – Neural Net Learning and Genetic Learning Parallel and Distributed AI: Psychological Modelling – Parallelism in Reasoning Systems – Distributed Reasoning Systems.

UNIT-IV: (9 Hrs)

Deep Learning Frameworks and AI Methodologies: Working – Framework – programming Languages – applications – optimization – fuzzy inference systems – artificial creativity – additional AI methodologies – glimpse into the future. Building DL network using MXNet, TensorFlow and Keras: Core components – MXNet, TensorFlow and Keras in action – Summary and Visualization.

UNIT-V: (9 Hrs)

Building and optimizer based on PSO and GA: Algorithm - implementation - variants - PSO and GA in action - Framework and tips. Building an Advanced DL system: CNN – RNN.

Alternative AI frameworks in DS: ELMs - CapsNets - Fuzzy logic and Fuzzy inference systems

Text Book:

1. Kevin Night, Elaine Rich, Nair B., "Artificial Intelligence (SIE)", McGraw Hill, 2008

Reference Books:

1. Russell, Stuart J., and Peter Norvig. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited, 2016.
2. Ivan Bratko, Prolog Programming for Artificial Intelligence (International Computer Science Series), Addison-Wesley Educational Publishers Inc; 4th edition, 2011.

Web References:

- http://www.aispace.org/index.html
- https://www.britannica.com/technology/artificial-intelligence
- https://www.sas.com/en_in/insights/analytics/what-is-artificial-intelligence.html

IMAGE ANALYTICS
(Course code : 23PDSE41)

SEMESTER– IV	HOURS – 4	CREDITS – 3

Objective:

To be able to understand and develop proficiency in applying spatial filtering techniques to enhance and manipulate digital images for improved clarity and feature extraction.

Outcomes:

At the end of the course Learners will be ability to:

CO1: Recall and explain the fundamental concepts of digital image processing, including image representation, sampling, quantization, and basic pixel operations. (K1)

CO2: Comprehend the mathematical tools used in digital image processing, such as vector and matrix operations, and apply them to perform basic image transformations and operations. (K2)

CO3: Apply spatial filtering techniques for various image enhancement tasks, including smoothing, sharpening, and edge detection, in both the spatial and frequency domains. (K3)

CO4: Analyze and compare different color models and transformations, and demonstrate an understanding of image and video segmentation, demonization, and enhancement techniques. (K4)

CO5: Evaluate various techniques for image and video compression, and critically assess their impact on quality, size, and processing speed in different applications. (K5)

CO6: Design and implement object detection, recognition, and tracking algorithms for images and videos, leveraging texture models, classification techniques, and object tracking methodologies. (K6)

UNIT-I: (10 Hrs)

Digital image representation- Visual Perception- Sampling and Quantization- Basic Relations between Pixels- Mathematical Tools Used in Digital Image Processing: Fundamental Operations –Vector and Matric Operations- Image Transforms (DFT, DCT,DWT, Hadamard).

UNIT-II: (10 Hrs)

Fundamentals of spatial filtering: spatial correlation and convolution-smoothing blurring-sharpening – edge detection –**Basics of filtering in the frequency domain:** smoothing-blurring- sharpening – Histograms and basic statistical models of image.

UNIT-III: (10 Hrs)

Colour models and Transformations – Image and Video segmentation-Image and video demonizing - Image and Video enhancement- Image and Video compression.

UNIT-IV: (10 Hrs)

Object detection and recognition in image and video – Texture models Image and Video classification models – Object tracking in Video.

UNIT-V: (10 Hrs)

Applications and Case studies – Industrial – Retail- Transportation & Travel – Remote sensing-Video Analytics in WSN: IoT Video Analytics Architectures.

Text Books:

1. R.C. Gonzalez and R.E. Woods." Digital Image Processing". 3rd Edition, AddisonWesley, 2007.
2. Caifeng Shan, Fatih Porikli, Tao Xiang, Shaogang Gong, "Video Analytics for Business Intelligence", Springer, 2012.

Reference Books:

- W. Härdle, M. Müller, S. Sperlich, A. Werwatz, "Nonparametric and Semiparametric Models", Springer, 2004.
- Rick Szelisk, "Computer Vision: Algorithms and Applications", Springer 2011.
- Jean-Yves Dufour, "Intelligent Video Surveillance Systems", Wiley, 2013.
- AsierPerallos,UnaiHernandez-Jayo,EnriqueOnieva, Ignacio Julio García Zuazola, "Intelligent Transport Systems: Technologies and Applications", Wiley, 2015.
- Basudeb Bhatta, "Analysis of Urban Growth and Sprawl from Remote Sensing Data", Springer, 2010.

CLOUD COMPUTING
(Course code: 23PDSE41)

SEMESTER– IV	HOURS – 4	CREDITS – 3

Objective:

To be able to gain expertise in cloud computing by understanding the basics concepts including models, virtualization, applications, and the security aspects.

Outcomes:

At the end of the course learners will be ability to

CO1: Explain the models, principles, and benefits of Cloud Computing (K2)
CO2: Summarize virtualization foundations to cater the needs of elasticity, portability, and resilience by cloud service providers. (K2)
CO3: Identify the applications of Cloud Computing (K1)
CO4: Analyze the security aspects of Cloud Computing (K4)
CO5: Evaluate the importance of message passing and map reduce in Cloud Computing (K5)
CO6: Create and deploy services in cloud infrastructure. (K6)

Unit I: **(15 Hrs)**

Cloud Computing: Introduction – History - Fundamentals of Cloud computing – characteristics - Advantages and Disadvantages- Comparison of traditional and cloud computing paradigms- Evaluating the impact and economics - Business drivers- Future of cloud. Services and Deployment model: Cloud deployment models - Cloud service models – Cloud infrastructure mechanisms - Cloud service management.

Unit II: **(14 Hrs)**

Cloud Computing Architecture: Cloud computing architecture - Design principle - Life cycle (CCLC) -Reference architecture - Load balancing approach - Mobile cloud computing (MCC) - Case study of oracle cloud management. Virtualization: Understanding - Adoption – Techniques – Working of Virtualization - Kernel-based virtual machine (KVM) – VMware – Virtual Box – Citrix - Types of virtualizations - Virtualization in cloud.

Unit III: **(14 Hrs)**

Service Oriented Architecture: Objectives - SOA foundation - Web services and SOA - SOA communication - SOA components - SOA Infrastructure - Need of SOA - Business Process Management (BPM) – Services of BPM. Cloud Computing Applications: Introduction - Google App Engine - Google Apps - Google Cloud Data store - Dropbox Cloud - Apple iCloud - Microsoft Windows Azure Cloud - Amazon Web Services.

Unit IV: **(14 Hrs)**

Cloud Security and Privacy: Cloud Security - Cloud CIA security model - Cloud computing security Architecture - Service provider security issues - Security issues in Virtualization - Data security in cloud – Data privacy risks - Business continuity and disaster recovery - Threats in cloud – Security techniques for threats - Cloud service level agreements (SLA): Components – Types - Cloud vendors - Quality of Cloud Services - Techniques – Migration - Trust management.

Unit V: **(15 Hrs)**

Cloud Computing Technologies: Cloud Computing Technologies - High performance Computing - Message Passing Interface (MPI) - MapReduce programming model -Dryad and Dryad LINQ -Eucalyptus cloud platform: Components – OpenNebula: Layers – Features – OpenStack: components - Benefits – The Apache Hadoop ecosystem. Adoption of Cloud

Computing: Factors affecting the adoption - Existing areas of application - Case studies - Certifications.

Text Book:

1. Kant Hiran, Kamal, Ruchi Doshi, Temitayo Fagbola, Mehul Mahrishi, Cloud Computing: Master the Concepts, Architecture and Applications with Real-world examples and Case studies, BPB Publishers, 2019.

Reference Books:

1. Ben Piper and David Clinton, AWS Certified Solutions Architect Study Guide: Associate SAA-C01 Exam, Googel Book, 2019.
2. Legorie Rajan Ps, Steven Porter, and Ted Hunter, Building Google Cloud Platform Solutions: Develop Scalable Applications from Scratch and Make Them Globally Available in Almost Any Language, Packt,2019.
3. Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi, Mastering Cloud Computing, Tata McGraw Hill Education Private Limited, 2013.

Web References:

- https://acloudguru.com
- https://www.cloudcomputing-news.net/
- https://cloudtweaks.com/

RECOMMENDATION SYSTEM
(Course code : 23PDSE41)

SEMESTER– IV	HOURS – 4	CREDITS – 3

Objective:

To understand the fundamentals of real time images, image transformations, detect edges and recognize objects in the image.

Outcomes:

At the end of the course Learners will be ability to:

CO1: Contrast the fundamental concepts and significance of recommender systems, including the various types of recommendation approaches and their applications in different domains. (K2)

CO2: Apply user-based and item-based collaborative filtering techniques to recommend items to users, and demonstrate proficiency in matrix factorization methods for building recommendation models. (K3)

CO3: Analyze the principles of content-based recommendation and effectively use TF-IDF and feature representation to create content-based recommenders. Evaluate the benefits and challenges of hybrid recommendation systems. (K4)

CO4: Evaluate the efficacy of advanced recommendation techniques such as matrix factorization techniques (SVD, PMF), factorization machines, and neural network approaches. (K5)

CO5: Design and develop evaluation methods for recommender systems, address scalability challenges in recommendations, and propose strategies to ensure privacy and ethics in recommender system design. (K6)

CO6: Recall and discuss future trends in recommender system research, staying updated with the evolving landscape of recommendation technologies. (K2)

UNIT I: (10 Hrs)

Introduction to Recommender Systems: Definition and Importance of Recommender Systems - Types of Recommendation Approaches - Challenges and Applications of Recommender Systems.

UNIT II: (10 Hrs)

Collaborative Filtering Techniques: User-Based Collaborative Filtering - Item-Based Collaborative Filtering - Matrix Factorization for Recommendations - Evaluation Metrics for Collaborative Filtering.

UNIT III: (10 Hrs)

Content-Based and Hybrid Approaches: Content-Based Recommendation Principles - TF-IDF and Feature Representation - Hybrid Recommender Systems - Combining Collaborative Filtering and Content-Based Methods.

UNIT IV: (10 Hrs)

Advanced Recommendation Techniques: Matrix Factorization Techniques (SVD, PMF, etc.) - Factorization Machines for Recommendations - Neural Network Approaches for Recommender Systems - Context-Aware and Session-Based Recommendations.

UNIT V: (10 Hrs)

Evaluation, Scalability, and Ethical Considerations: Evaluation Methods for Recommender Systems - Scalability Challenges and Solutions - Privacy and Ethical Considerations in Recommender Systems - Future Trends in Recommender System Research.

Text Book:

1. Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich, Recommender Systems: An Introduction, Cambridge University Press, 2010.

Reference Books:

- Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor, Recommender Systems Handbook, Springer, 2010.
- Toby Segaran, "Programming Collective Intelligence: Building Smart Web 2.0 Applications", O'Reilly Media, 2007.

DEEP LEARNING LAB
(Course Code: 23PDSS41)

SEMESTER– IV	Hours – 4	Credits –2

Objective:

To explore the principles of deep learning techniques involved in simulating an virtual environment and apply the same in the field of Data Science.

Outcomes:

At the end of the course Learners will be ability to

CO1: Recall and explain key neural network concepts, including layers, activation functions, loss functions, and optimizers. (K1)

CO2: Apply various deep learning techniques for building neural networks, designing CNN architectures, using pre-trained models, creating RNN models, and constructing GANs for image generation. (K3)

CO3: Analyze model performance metrics, compare outcomes of transfer learning versus conventional training, and assess the influence of architectural choices on results. (K4)

CO4: Design comprehensive solutions by integrating knowledge from different units to address complex tasks like image analysis, text generation, and generative modeling. (K6)

CO5: Evaluate the ethical implications of AI, examine potential biases in datasets, and assess the broader societal impact of AI-generated content and automation. (K5)

CO6: Synthesize their learning by conceptualizing and implementing original applications that combine neural networks, CNNs, transfer learning, RNNs, and GANs to solve real-world challenges. (K6)

Unit I: **(8 hrs)**

Building a Simple Neural Network with Keras: Import the necessary libraries: numpy, keras - Create a simple dataset (e.g., XOR problem) using numpy - Build a sequential model with Keras containing an input layer, hidden layer(s), and an output layer - Compile the model with an appropriate loss function and optimizer - Train the model on the dataset - Evaluate the model's performance and visualize the results.

Unit II: **(8 hrs)**

Convolutional Neural Network (CNN) for Image Classification: Import required libraries:tensorflow, keras - Load and preprocess a dataset (e.g., CIFAR-10) using Keras preprocessing functions - Design a CNN architecture with convolutional layers, pooling layers, and fully connected layers - Compile the model with appropriate loss and optimizer - Train the CNN on the dataset and monitor the training process - Evaluate the model's accuracy on the test dataset and analyze misclassified samples.

Unit III: **(8 hrs)**

Transfer Learning with Pre-trained Models: Import libraries: tensorflow, keras - Load a pre-trained model (e.g., VGG16, ResNet50) from Keras applications - Replace the classification head of the pre-trained model with a new one for a different task - Fine-tune the model on a new dataset (e.g., flowers, dogs vs. cats) - Train the modified model and evaluate its performance - Compare the results with and without transfer learning.

Unit IV: **(8 hrs)**

Recurrent Neural Networks (RNNs) for Text Generation: Import libraries: tensorflow, keras. - Prepare a dataset of text sequences (e.g., song lyrics, movie scripts) - Tokenize the text and create sequences for input and target - Build an RNN model using LSTM layers - Train the model to predict the next word in a sequence - Generate new text samples using the trained model.

Unit V: **(8 hrs)**

Generative Adversarial Networks (GANs): Import libraries: tensorflow, keras - Build a GAN architecture with a generator and a discriminator network - Train the GAN to generate realistic images (e.g., handwritten digits) - Visualize the generated images during the training process - Experiment with different loss functions and architectures to improve image quality.

Text Books:

1. Francois Chollet, Deep Learning with Python, Manning, First Edition, 2017.
2. Seth Weidman, Deep Learning from Scratch: Building with Python from First Principles, O′Reilly, 2019.

Reference Books:

- Sudharsan Ravichandiran, Hands-On Deep Learning Algorithms with Python, Packt Publishing Limited, 2019.
- S Lovelyn Rose, L Ashok Kumar, D Karthika Renuka, Deep Learning Using Python, Wiley, 2019.

(EXTENSION ACTIVITY]
STAND

SEMESTER– IV	**CREDITS – 1**

ADD ON COURSES

EXCEL (Basic Level)
(Course Code: 23PDSAO1)

SEMESTER– I	HOURS – 35	CREDITS – 3

Objective:

To be able efficiently organize large amounts of data and create tables, format cells, and use features like sorting and filtering to manage your data effectively.

Outcomes:

At the end of the course Learners will be ability to

CO1: Apply the Basic functions in Excel. (K2)
CO2: Create a simple worksheet in Excel (K6)
CO3: Formulate the Basic formulas and functions in Excel. (K6)
CO4: Apply various functions and formulas for different data. (K3)
CO5: Generate a simple chart in Excel. (K6)
CO6: Compare the data using Mathematical functions. (K4)

UNIT I: **(7 hrs)**

Introduction to Excel: Understanding Workbooks- Navigating-File- Typing Text Numbers into A Worksheet- Typing Simple Formulas in A Worksheet- Inserting and Deleting Worksheets.

UNIT II: **(7 hrs)**

Changing Worksheet Tab Colours- Freezing Rows and Columns- Selecting Ranges, Rows And Columns- Changing Fonts- Understanding Borders-Applying A Border To A Range-Wrapping And Merging Text.

UNITIII: **(7 hrs)**

Building Basic Formulas - Formula building methods - Editing formulas - Building formulas with computational operators.

UNIT IV: **(7 hrs)**

Building Basic Formulas-Creating Formulas.: Using Formulas, Formula Functions – Sum, Average, if, Count, max, min, Proper, Upper, Lower. Using AutoSum - Logical Functions - Understanding Error Values.

UNIT V: **(7 hrs)**

Trim Spreadsheet Charts: Creating Charts, Different types of charts, Formatting Chart Objects, Changing the Chart Type, Showing and Hiding the Legend, Showing and Hiding the Data Table

Text Book:

1. Greg Harvey, Microsoft Excel 2019 All in One : John Wiley & Sons, 2019.

Reference Books:

1. Joseph Schmuller, Statistics Analysis with Excel, John Wiley & Sons, 3rd Edition.
2. Stephen L. Nelson, Microsoft Data Analysis, John Wiley & Sons, 3rd Edition.

Web References:

- https://www.analyticsvidhya.com/
- https://www.simplilearn.com/tutorials/excel-tutorial/data-analysis-excel
- https://www.coursera.org/learn/excel-basics-data-analysis-ibm

EXCEL (INTERMEDIATE LEVEL)
(Course Code: 23PDSAO2)

SEMESTER– I	HOURS – 35	CREDITS – 3

Objective:

It allows you to efficiently enter and organize large amounts of data. You can create tables, format cells, and use features like sorting and filtering to manage your data effectively.

Outcomes:

At the end of the course Learners will be ability to

- **CO1:** Understand the macros in Excel (K2)
- **CO2:** Apply the Navigation concepts in Excel. (K3)
- **CO3:** Leverage sampling techniques from data sets using the Data Analysis command (K3)
- **CO4:** Design a Worksheet using Graphics tools. (K6)
- **CO5:** Apply Financial tools to transform data into useful information. (K3)
- **CO6:** Generate a worksheet using Statistical concepts. (K6)

UNIT I: (7 hrs)

Protecting Workbooks and Worksheet Data- Assigning a password to open from the Info screen- Assigning a password to unprotect the sheet- Removing protection from a worksheet- Enabling cell range editing by certain users-Making Macros-Recording Macros-Running Macros-Deleting Macros.

UNIT II: (7 hrs)

Using Hyperlinks- Adding hyperlinks- Editing hyperlinks- Removing a hyperlink- Copying and moving a hyperlink- Using the HYPERLINK Function.

UNIT III (7 hrs)

Charting Worksheet Data - Inserting specific chart types from the Ribbon- Inserting charts with the Quick Analysis tool- Customizing chart elements from the Format tab- Adding a data table to a chart- Editing the chart titles- Formatting elements of a chart- Adding Sparkline Graphics to a Worksheet- Adding Infographics to a Worksheet.

UNIT IV (7 hrs)

Advanced formulas- Financial Functions- The PV, NPV, and FV Functions- The PMT Function- Depreciation Functions- Math & Trig Functions- Rounding off numbers- POWER and SQRT- SUMIF function.

UNIT V (7 hrs)

Statistical Functions- AVERAGE, MAX, and MIN- Counting cells- Conditional counting- Lookup functions- VLOOKUP and HLOOKUP- Performing a two-way lookup- Match and Index function. Using the IS functions.

Text Books:

1. Greg Harvey, Microsoft Excel 2019 All in One : John Wiley & Sons, 2019.

Reference Books:

1. Joseph Schmuller, Statistics Anaysiswith Excel, John Wiley & Sons, 3rd Edition.
2. Stephen L.Nelson, Microsoft Data Analysis, John Wiley & Sons, 3rd Edition.

Web References:

1. https://www.analyticsvidhya.com/
2. https://www.simplilearn.com/tutorials/excel-tutorial/data-analysis-excel
3. https://www.coursera.org/learn/excel-basics-data-analysis-ibm

R (Basic Level)
(Course Code: 23PDSAO3)

SEMESTER– II	HOURS – 35	CREDITS – 3

Objective:

To be able to develop proficiency in using R for data analysis, manipulation, and visualization, apply statistical techniques, build predictive models, evaluate results, and implement reproducible data analysis workflows using R Markdown and Git.

Outcomes:

At the end of the course the students must be able to

CO1: Demonstrate proficiency in using R programming language for data analysis, manipulation, and visualization. (K1)

CO2: Apply fundamental statistical techniques to explore and summarize data effectively. (K2)

CO3: Analyze the packages in R for efficient data manipulation and transformation. (K3)

CO4: Build predictive models using linear regression, logistic regression, and other statistical modeling techniques in R. (K4)

CO5: Evaluate and interpret the results of statistical models, including assessing model fit and variable importance. (K5)

CO6: Develop reproducible data analysis workflows using R Markdown and version control tools like Git, ensuring transparent and collaborative data science practices. (K6)

Unit I: **(7 hrs)**

Introduction to R: What is R - Why R – Installing – Environment - data structure - Variables – Scalars – Vectors.

Unit II: **(7 hrs)**

Introduction: Matrices – List. R Programming: While - If structure - For - Arithmetic operations.

Unit III: **(7 hrs)**

Data frames - attach - detach - Importing data - Tabular Data - CSV - data from excel - Accessing database – Loading – Writing to files.

Unit IV: **(7 hrs)**

Manipulating Data: Selecting rows/observations - Selecting columns/fields - Merging data - Relabelling the column names - Converting variable types - Data sorting - Data aggregation.

Unit V: **(7 hrs)**

Using functions in R: Mathematical Functions - String Functions - local and global variable - Box plot - Line chart - Pie graph.

Text Books

1. Roger D Peng, R programming for Data science, Lulu.com, 2012.

Reference Books

- Michael J. Crawley, The R Book, 2nd Edition, 2012.

R (Intermediate Level)
(Course Code: 23PDSAO2)

SEMESTER– II	HOURS – 35	CREDITS – 3

Objective:

To be able to develop proficiency in applying statistical techniques, build predictive models, evaluate results, and implement reproducible data analysis workflows using R Markdown and Git.

Outcomes:

At the end of the course the students must be able to

CO1: Describe the R programming basic for data analysis and manipulation. (K2)
CO2: Apply skills in exploratory data analysis and visualization using R. (K3)
CO3: Apply various supervised and unsupervised machine learning algorithms in R.(K4)
CO4: Compare the techniques for feature engineering, selection, and preprocessing to enhance the performance of predictive models. (K5)
CO5: Demonstrate data cleaning and preprocessing techniques for effective data analysis. (K4)
CO6: Develop procedures to solve problems related to time series analysis, text mining, and ethical considerations in data science applications. (K6)

Unit I: **(7 hrs)**

Introduction to Data Science with R: Understanding the data science process - R as a tool for data science: advantages and capabilities - Installing R and RStudio - RStudio interface and basic operations - Introduction to R packages for data science - Exploring and manipulating data using dplyr and tidyr - Data visualization with ggplot2 - Introduction to the caret package for machine learning in R.

Unit II: **(7 hrs)**

Data Preparation and Cleaning: Overview of data preprocessing tasks - Dealing with missing data: imputation techniques - Handling outliers and anomalies - Data transformation and normalization. Feature engineering: creating new variables and derived features - Exploratory data analysis (EDA) using visualizations - Sampling techniques for large datasets - Splitting data into training and testing sets.

Unit III: **(7 hrs)**

Predictive Modeling and Machine Learning: supervised learning - classification and regression - Model training and evaluation using the caret package - Decision trees and random forests - Support Vector Machines - Ensemble methods: bagging and boosting - Evaluation metrics for classification and regression models - Feature selection and variable importance - Model interpretation and understanding.

Unit IV: **(7 hrs)**

Advanced Techniques in Data Science: Handling large datasets with data, Table and dplyr - Text mining and natural language processing - Introduction to deep learning with the keras package - Time series analysis and forecasting - Dimensionality reduction techniques: PCA and t-SNE. Unsupervised learning: clustering algorithms - Anomaly detection in data - Model deployment and operationalization.

Unit V: **(7 hrs)**

Case studies: finance – healthcare – marketing. Data science project lifecycle: problem formulation - data acquisition – modeling – evaluation – deployment. Building end-to-end data science workflows in R - Best practices for reproducible and scalable data science

projects - Ethical considerations in data science - Communicating and presenting data science results - Practical tips and resources for further learning.

Text Books:

1. Nina Zumel and John Mount,Practical Data Science with R, Manning, 2nd Edition, 2019.

Reference Books:

1. Roger D Peng, R programming for Data science, Lulu.com, 2012.
2. Michael J. Crawley, The R Book, 2nd Edition, 2012.

PYTHON FOR DATA SCIENCE
(Course code:23PDSAO5)

SEMESTER– III	HOURS -35	CREDITS – 5

Objective:

Ability to handle various stages of the data science lifecycle, from data acquisition to model deployment.

Outcomes:

By the end of the course, students should be able to:

CO1: Understand the fundamentals of Python programming and its relevance to data science. (K2)
CO2: Manipulate and clean data using Python libraries. (K3)
CO3: Visualize data effectively to gain insights. (K3)
CO4: Perform basic statistical analysis and hypothesis testing. (K1)
CO5: Build and evaluate predictive models. (K6)
CO6: Interpret and communicate results to non-technical stakeholders. (K5)

UNIT I: (7hrs)

Introduction to Python for Data Science: Introduction to data science and its applications-Setting up the Python environment (Python installation, Jupyter Notebooks)-Python basics: data types, variables, operators, loops, and conditionals-Data Manipulation with Pandas-Introduction to Pandas library for data manipulation-Loading and exploring data with Pandas Data Frames-Data cleaning and preprocessing-Combining and merging datasets.

UNIT II: (7hrs)

Data Visualization with Matplotlib and Seaborn: Introduction to data visualization Principles-Creating basic plots with Matplotlib-Advanced visualization techniques with Seaborn-Customizing plots for effective communication- Exploratory Data Analysis (EDA)-The role of EDA in data science-Statistical summaries and visualizations for EDA-Handling missing data and outliers-Extracting insights from data distribution analysis.

Unit III: (7hrs)

Basic Statistical Analysis and Hypothesis Testing: Understanding basic statistical concepts - mean - median - standard deviation - Introduction to probability distributions-Hypothesis testing and p-values-Applying t-tests and chi-squared tests.

UNIT IV: (7hrs)

Introduction to Machine Learning: Overview of machine learning concepts-Splitting data into training and testing sets-Supervised vs. unsupervised learning-Linear regression for predictive modelling.

UNIT V: (7hrs)

Model Evaluation and Cross-Validation: Evaluating model performance metrics (accuracy, precision, recall, F1-score, etc.)-Overfitting and underfitting-Cross-validation techniques-Introduction to hyperparameter tuning.

Text Books:

1. Jake VanderPlas, "Python Data Science Handbook",O'Reilly Publication, First Edition,2016.
2. Peter Bruce and Andrew Bruce," Practical Statistics for Data Scientists", O'Reilly Publication, First Edition,2017.

Reference Books:

1. 1.Luca Massaron John Paul Muller,"Python for Data Science for Dummies", Wiley Publication, Second Edition,2019.
2. Andreas C. Müller and Sarah Guido, "Introduction to Machine Learning with Python", O'Reilly Publication, First Edition, 2016.

Web Resources:

- 1.https://www.udemy.com/course/

NATURAL LANGUAGE PROCESSING LAB
(Course code : 23PDSAO6)

SEMESTER– III	HOURS – 3	CREDITS – 3

Objective:

To explore the basics of NLP which leads to the development of advanced NLP tools and solving practical problems in the chosen domain.

Outcome:

At the end of the course the students will be able to

CO1: Summarize the fundamentals of Natural Language Processing. (K2)
CO2: Apply the NLP techniques for word and syntactic analysis. (K3)
CO3: Analyze the natural language text. (K4)
CO4: Evaluate the tools and methods for understanding semantics of sentences and pragmatics. (K5)
CO5: Develop an innovative application using NLP components. (K6)
CO6: Evaluate an application using NLP components. (K5)

UNIT-I: (7 hrs)

INTRODUCTION TO NLP: Speech and Language Processing – Ambiguity - Models and Algorithms- Language, Thought, and Understanding - State of the Art - Applications – Basic NLP.

UNIT-II: (7 hrs)

WORD ANALYSIS: Regular Expressions - Words & Transducers- Survey of English Morphology- Finite-State Morphological Parsing - Word and Sentence Tokenization- N-grams- Counting Words in Corpora- Simple (Unsmoothed) N-grams- Training and Test Sets- Part-of-Speech Tagging- English Word Classes- Tagsets for English- Part-of-Speech Tagging- Rule-Based Part-of-Speech Tagging- Evaluation and Error Analysis.

UNIT-III: (7 hrs)

SYNTACTIC ANALYSIS: Formal Grammars of English- Constituency- Context-Free Grammars- Grammar Rules for English - Treebanks - Finite-State and Context-Free Grammars - Dependency Grammars - Parsing with Context-Free Grammars - Parsing as Search- Dynamic Programming Parsing Methods- Statistical Parsing- Probabilistic Context-Free Grammars- Probabilistic CKY Parsing of PCFGs- Learning PCFG Rule Probabilities.

UNIT-IV: (7 hrs)

SEMANTICS AND PRAGMATICS: Computational Desiderata for Representations- First-Order Logic, Computational Semantics –Syntax Driven Semantic analysis, Semantic attachments Semantic Attachments for a Fragment of English, Lexical Semantics- Word Senses, Relations between Senses, WordNet: A Database of Lexical Relations- Event Participants: Semantic Roles and Selectional Restriction.

UNIT-V: (8 hrs)

APPLICATIONS: Applications - Information Extraction, Question Answering and Summarization, Dialogue and Conversational Agents.

List of Experiments:

1. Word Analysis
2. Word Generation
3. Morphology
4. N-Grams
5. N-Grams Smoothing

6. POS Tagging: Hidden Markov Model
7. POS Tagging: Viterbi Decoding
8. Building POS Tagger
9. Chunking
10. Building Chunker
11. Analyse words with WordNet with different operations
12. Use Word Embedding techniques

Text Books:

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.Chapter 1,2,3,4,5,12,13,14,17,18,19,22.23.24

2. Patel, Ankur A., and Ajay UppiliArasanipalai. Applied Natural Language Processing in the Enterprise. " O'Reilly Media, Inc.", 2021.
Chapter1.

Reference Books:

1. Breck Baldwin, Language Processing with Java and Ling Pipe Cookbook, Atlantic Publisher, 2015.
2. Richard M Reese, Natural Language Processing with Java‖, O_Reilly Media, 2015.
3. Nitin Indurkhya and Fred J. Damerau, Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.
4. Tanveer Siddiqui, U.S. Tiwary, Natural Language Processing and Information Retrieval‖, Oxford University Press, 2008.

Web References:

- https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/what-is-natural-language-processing-nlp
- 2.https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1
- 3.https://www.oracle.com/in/artificial-intelligence/what-is-natural-language-processing/

Power BI
(Course Code: 23PDSAO7)

SEMESTER– IV	HOURS – 35	CREDITS – 3

Objective:

To be able to Navigate the Power BI interface and perform basic data transformations and build effective data models using relationships, hierarchies, and calculated columns.

Outcomes:

At the end of the course the students must be able to

CO1: Describe the functions and use of Power BI. (K1)
CO2: Explain designing and building interactive visualizations using different chart types. (K2)
CO3: Apply concepts learned in excel to create calculated columns and measures to perform advanced calculations. (K3)
CO4: Combine multiple data sources using relationships to build robust data models. (K4)
CO5: Publish and share reports securely with colleagues and stakeholders. (K5)
CO6: Apply data visualization principles to effectively present insights. (K6)

Unit I: **(7 hrs)**

Introduction to Power BI – Working with data – Importing from flat files, excel files, other sources – Power Pivot data types – Column operations - Table relationship.

Unit II: **(7 hrs)**

Power Pivot data analysis – PivotTable and PivotChart – Slicers – Dashboard Implementation – Dates, hierarchies, and perspectives.

Unit III: **(7 hrs)**

Data Analysis Expressions – Introduction to Power Query – Introduction to Power View – Power View visualizations.

Unit IV: **(7 hrs)**

Power View filtering options – Introduction to Power Map – Preparing geospatial data.

Unit V: **(7 hrs)**

Publish from Power BI desktop – Publish Dashboard to Web.

Text Books

1. Microsoft Power BI cookbook, Brett Powell, Packt Publishing Limited, 2nd edition, 2017.

Reference Books

- Greg Deckler, Brett Powell, Microsoft Power BI Cookbook: Gain expertise in Power BI, Packt Publishing Limited, 2nd Edition, 2021.

TABLEAU
(Course Code: 23PDSAO8)

SEMESTER– IV	HOURS –40	CREDITS – 3

Objective:

To explore on the tableau tool used for data visualization and analytics.

Outcomes:

At the end of the course the students will be able to

CO1: Recite the basic elements of data visualization (K1)
CO2: Differentiate the basic principles of data preprocessing in tableau (K2)
CO3: Illustrate visualization functions in Tableau (K3)
CO4: Outline the use of shapes, colors, text, and layout appropriately (K4)
CO5: Evaluate the use of Multiple sources to customize the story in tableau (K5)
CO6: Create ad-hoc reports, data visualizations, and dashboards using Tableau Desktop (K6)

List of Practical's:

1. Create a chart after filtering and sorting the data in the dataset.
2. Using the dataset assemble a dashboard layout and apply dashboard filters
3. Use tableau to formulate simple calculations
4. Use tableau to create a calculator for table calculations
5. Examine interactivity with text and visual tooltips
6. Interactivity with actions such as filter, highlight and URL
7. Create more advanced chart types using multiple source tables
8. Create a data story in Tableau

Text Books:

1. Joshua N. Milligan, "Learning Tableau 2022", 5th Edition, Pockt Publishing, 2022.

Reference Books:

- Ryan Sleeper, "Practical Tableau", O'Reilly, 2018.

Value Added Courses

EXCEL FOR STATISTICAL AND DATA ANALYTICS
(Course Code: 23PDSVA1)

SEMESTER– I	HOURS –40	CREDITS – 3

Objective:

The objectives of using Excel in data analysis revolve around organizing, cleaning, summarizing, visualizing, analyzing, and modeling data to gain insights and make informed decisions.

Outcomes:

At the end of the course the students will be able to:

CO1: Apply Business Intelligence tools to transform data into useful information. (K3)
CO2: Compute the Statistical functions using various data. (K3)
CO3: Test the data using Statistical function. (K5)
CO4: Apply various formulas for data analysis. (K6)
CO5: Apply the Data base function in Excel. (K6)
CO6: Analyse the data using regression functions. (K4)

UNIT-I **(8 hrs)**

Statistical Functions-Calculation of Mean, Median and Mode-Standard deviation function-Variance – Quartiles – Correlation – Percentile – Max – Min.

UNIT-II **(8 hrs)**

Tunning in frequency-Histogram-Normal Distribution-T-Distribution-Binomial Distribution-Binomial probability distribution-Negative Binomial Distribution-Cumulative Binomial Distribution-Hyper Geometric Distribution.

UNIT-III **(8 hrs)**

Chi-square Distribution – Righttailed Chi-Square Distribution – Lefttailed Chi-Square Distribution – Chi-Square Test – Hypothesis Testing – F-Test –Distribution.

Unit-IV **(8 hrs)**

Regression Analysis-FORECAST LINEAR function-FORECAST ETS-Slope of regression line-exponential regression-Exponential Growth-Correlation.

UNIT-V **(8 hrs)**

Introducing Excel Data Bases- Satellite Data Base-The Criteria Range-Data Base Function-DCOUNT-DCOUNTA-Arithmetic function-DMAX, DMIN, DSUM, DPRODUCT - Statistics function-DAVERAGE-DVAR-DVARO-DSTDEV-Pivot Tables.

Text Books:

1. Greg Harvey, Microsoft Excel 2019 All in One: John Wiley & Sons, Inc.2019 edition
2. Joseph Schmuller,StatisticsAnalysiswith Excel, John Wiley & Sons, 3rd Edition.
3. Stephen L.Nelson, Microsoft Data Analysis, John Wiley & Sons, 3rd Edition.

Web References:

1. https://www.analyticsvidhya.com/
2. https://www.simplilearn.com/tutorials/excel-tutorial/data-analysis-excel
3. https://www.coursera.org/learn/excel-basics-data-analysis-ibm

DATABASE LAB FOR DATA SCIENCE (Basic Level)
(Course Code: 23PDSVA2)

SEMESTER– II	HOURS – 35	CREDITS – 3

Objective:

The objectives of using Excel in data analysis revolve around organizing, cleaning, summarizing, visualizing, analyzing, and modeling data to gain insights and make informed decisions.

Outcomes:

At the end of the course the students must be able to:

CO1: Associate the basic concepts of Database and Oracle. (K3)
CO2: Explain the Basic concepts of ER model and DDL commands (K2)
CO3: Apply the Queries to manipulate the data in Oracle. (K3)
CO4: Analyze and Understand and the concept to implement the Join and set operation in Oracle. (K4)
CO5: Create a PL/SQL program to evaluate the SQL queries in Oracle(K6)
CO6: Evaluate a PL/SQL program using Control Statements. (K6)

Unit I: (12 Hours)

RELATIONAL DATABASE AND DATA BASE DESIGN: Purpose of Database systems-View of Data-Database Languages-Relational databases—Structure of relational databases-Data Base Architecture-Data Base users and Administrators- Overview of design process-E-R Model-E-R Diagram.

Unit II: (12 Hours)

DATA DEFINTION LANGUAGE: Data Types-Constraints-Creating an Oracle Table-Displaying Table Information-Altering Existing Table-Dropping Table-Renaming Table-Truncating Table-Spooling.

Unit III: (12 Hours)

DATA MANIPULATION AND RETRIEVAL: Data manipulation Language—Adding a New Row/Record-Updating Existing Rows/Records-Deleting Existing Rows/Records - Retrieving Data from a table-Arithmetic Operators-Group by Having Clause.

Unit IV: (12 Hours)

Introduction to NoSQL Databases: Overview of NoSQL Databases-Comparison with relational databases-Types of NoSQL databases: key-value stores-document stores- columnar databases-graph databases-Use cases and advantages of NoSQL databases.

Unit V: (12 Hours)

Key-Value Stores: Understanding key-value stores-Popular key-value store databases (e.g., Redis, Riak)- Data model and key-value pair storage-Querying and indexing in key-value stores - Caching strategies using key-value stores.

Text Books:

1. Abraham Silberschatz, Hendry F.Korth, S.Sundarsan, “Data Base System Concepts”, Sixth edition,McGraw Hill,2011.
2. Lemahieu, W., Broucke, S.Vanden, Baesens, B. (2018) Principles of database management:The Practical Guide to storing, managing, and analyzing big and small data. Cambridge, UnitedKingdom: Cambridge University Press.
3. Sadalage, P., Fowler,NoSQL Distilled: A Brief Guide to the Emerging World of PolyglotPersistence,Wiley Publications,1st Edition,2022

Reference Books:

1. Alexix Leon and Mathews Leon, Fundamentals of Data Base Management Systems,Vijay Nicole Imprints,2010.
2. SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis Renee M. P. Teate.
3. SQL for Data Science: Data cleaning, wrangling and analytics with relational databases, Antonio Badia
4. Guy Harrison, Next Generation Database: NoSQL and big data, Apress, 2018.

Web Resources:

- http://www.w3schools.com
- http://www.tuorialspoint.com
- http://www.javapoint.com

DATA BASE LAB FOR DATA SCIENCE (Advanced Level)
(Course code : 23PDSVA3)

SEMESTER– III	HOURS – 40	CREDITS – 3

Objective:

To provide fundamentals of database design, modeling systems, data storage, world of data warehousing and NoSQL.

Outcomes:

At the end of the course the students must be able to:

CO1: Discuss the importance of relational data modeling and conceptual Modelling (K2)
CO2: Experiment with various database and compose effective queries (K3)
CO3: Analyze the process of OLAP system construction (K4)
CO4: Evaluate the use of NOSQL and its approach to the database (K5)
CO5: Develop applications using Relational and NoSQL databases (K6)
CO6: Create a database using NoSQL. (K6)

UNIT I: (14hrs)

Fundamental Concepts of Database Management: Applications of Database Technology - File versus Database Approach to Data Management - Advantages of Database Systems and Database Management - Architecture and Categorization of DBMSs. **Conceptual Data Modelling using the ER Model and UML Class Diagram:** Phases of Database Design - The Entity Relationship Model - UML Class Diagram.

UNIT II: (14hrs)

Types of Database Systems: Legacy Databases - Relational Databases: The Relational Model – Normalization. **Relational Databases:** Structured Query Language - SQL Data Definition Language - SQL Data Manipulation Language.
Lab: SQL DDL and DML

UNIT III: (8 hrs)

Data Warehousing and Business Intelligence: Operational versus Tactical/Strategic Decision-Making - Data Warehouse Definition - Data Warehouse Schemas - Extraction, Transformation, and Loading (ETL) Process - Data Marts - Virtual Data Warehouses and Virtual Data Marts - Operational Data Store - Data Warehouses vs Data Lakes - Business Intelligence. **NO SQL:** History - Four Types - Value of Relational Databases – Concurrency – Integration - Impedance Mismatch - Attack of the Clusters - Key Points Comparison - Mongo DB - Cassandra, HBASE, Neo4j use and deployment, Application, RDBMS approach, Challenges NoSQL approach, Key-Value and Document Data Models, Column-Family Stores, AggregateOriented Databases. sharding, MapReduce on databases. Distribution Models, Single Server, Sharding, Master-Slave Replication, Peer-to-Peer replication, Combining Sharding and Replication.

UNIT IV: (8 hrs)

Key Value Data Stores: NoSQL Key/Value databases using MongoDB, Document Databases, Document oriented Database Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Event Logging, Content Management Systems, Blogging Platforms, Web Analytics or Real-Time Analytics, E-Commerce Applications, Complex Transactions Spanning Different Operations, Queries against Varying Aggregate Structure.
Lab: Key-value databases, Replica of existing database,Backup of existing database, Restore database from the backup
Demonstration: Connecting python with mongodb and inserting, retrieving, updating,anddeleting.

UNIT V: (8 hrs)

Document Oriented Database: Column- oriented NoSQL databases using Apache HBASE, Column-oriented NoSQL databases using Apache Cassandra, Architecture of HBASE, Column-Family Data Store Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Event Logging, Content Management Systems, Blogging Platforms, Counters, Expiring Usage. **Data Modeling with Graph:** Comparison of Relational and Graph Modeling, Property Graph Model Graph Analytics: Link analysis algorithm- Web as a graph, Page Rank Markov chain, page rank computation, Topic specific page rank Page Ranking Computation techniques iterative processing, Random walk distribution Querying Graphs.

Lab: Implement with column-family stores(cassandra), Graph databases (neo4j), Aggregate function, Push and addtoset expression, First and last expression.

Text Books:

1. Lemahieu, W., Broucke, S. Vanden, Baesens, B. Principles of database management: The Practical Guide to storing, managing, and analyzing big and small data. Cambridge, United Kingdom: Cambridge University Press, 2018.
2. Sadalage, P. Fowler,NoSQL Distilled: A Brief Guide to the Emerging World of PolyglotPersistence,Wiley Publications,1st Edition,2022.

Reference Books:

1. Renee M. P. Teate,SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis, Kindle Edition, 2021.
2. Antonio Badia,SQL for Data Science: Data cleaning, wrangling and analytics with relational databases, Springer, 2020.
3. Guy Harrison, Next Generation Database: NoSQL and big data, Apress, 2015.

Web References:

- https://www.geeksforgeeks.org/introduction-to-nosql/

CLOUD COMPUTING LAB
(Course Code: 23PDSVA4)

SEMESTER– IV	Hours – 4	Credits – 3

Objective:
To explore the principles of cloud computing and techniques involved in simulating an virtual environment and apply the same in the field of Data Science.

Outcomes:
CO1: Examine virtualization using simulators and cloud platforms. (K5)
CO2: Creating cloud and virtual computing environment using simulators for executing algorithms as well as launching applications. (K6)
CO3: Develop applications using Google App Engine. (K6)
CO4: Experiment the use of Hadoop to handle data by developing programs. (K5)
CO5: Explore on deploying and developing applications in Microsoft Azure. (K4)
CO6: Develop an application using HADOOP. (K6)

List of Practicals:
1. Install Virtual box / VMware Workstation with different flavors of Linux or windows OS on top of Windows OS.
2. Install a C compiler in the virtual machine created using virtual box and execute Simple Programs
3. Find a procedure to transfer the files from one virtual machine to another virtual machine.
4. Simulate a cloud scenario using CloudSim and run a scheduling algorithm that is not present in CloudSim.
5. Create hello world app and other simple web applications using python/java using Cloud Platforms.
6. Deploy web applications in Cloud Platforms.
7. Install Hadoop single node cluster and develop a simple wordcount application to count the number of occurrences of a word in each set of input.
8. Develop a Hadoop application to process given data and produce results such as finding the year of maximum usage, year of minimum usage.
9. Develop a Windows Azure Hello World application.
10. Use Azure Web apps for hosting web sites and configure web app deployment settings.

Certificate Courses

JAVA PROGRAMMING LAB (Advanced Level)

(Course Code: 23PDSCC1)

SEMESTER– I	HOURS –30	CREDITS – 3

Objective:

To be able to apply the Java programming knowledge for exploring the basics of object-oriented concepts and develop apps for solving real life problems using graphical user interface/ command line interface-based packages and classes.

Outcomes:

At the end of the course the students must be able to:

CO1: Differentiate the object-oriented concepts of java (K2)
CO2: Apply selection and looping constructs, appropriate for solving problems (K3)
CO3: Examine and implement the concepts of object-orientation using Java (K3)
CO4: Analyze the use of GUI components including classes and functions (K4)
CO5: Compare use different packages and classes using java (K5)
CO6: Develop application for solving real time issues using java. (K6)

Unit I: **(6 hrs)**

Java Fundamentals - Features of Java - OOPs concepts - Java virtual machine - Reflection byte codes - Byte code interpretation - Data types – variable – arrays – expressions – operators – control structures - Objects and classes.

Unit II: **(6 hrs)**

Java Classes - Abstract classes - Static classes - Inner classes – Packages - Wrapper classes – Interfaces – This – Super - Access control.

Unit III: **(6 hrs)**

Exception handling: Exception as objects - Exception hierarchy - Try catch finally - Throw, throws. **IO package:** Input streams - Output streams - Object serialization – Deserialization - Files - Filter - pipe streams.

Unit IV: **(6 hrs)**

Multi-threading: Thread Life cycle - Multi threading - Simple thread program - Thread synchronization. **GUI**: AWT programming - Layout and component managers - Event handling - Applet class - Passing parameters embedding in HTML.

Unit V: **(6 hrs)**

Swing components – JApplet – Jbutton – Jframe - swing programs - JDBC architecture - connection interface - Working with statements - SQL statements - Result Set – RMI – Servlets.

Text Books

1. E. Balaguruswamy, Programming with Java A Primer, Tata McGraw Hill, 3rd Ed, 2006.
2. John P. Flynt, Java Programming for the Absolute Beginner, 2nd edition, 2006.

Reference Books

- Cay Horstmann, Big Java, Wiley India Edition, 2nd edition, 2016.
- Ivan Bayross, Vaishali Shah, Sharanam Shah, Cynthia Bayross, Java server programming, 2nd edition, 2008.

WEB PROGRAMMING
(Course Code: 23PDSCC2)

SEMESTER– I	HOURS –30	CREDITS –3

Objective:
To introduce students about web application and state management.

Outcomes:
At the end of the course the students will be able to
CO1: Design a static web page and relate the basic structure of PHP (K6)
CO2: Demonstrate the syntax and semantics of PHP programs. (K4)
CO3: Apply functions of PHP in designing web sites (K4)
CO4: Evaluate the concepts of PHP frame works. (K4)
CO5: Develop websites using built in-functions with PHP/MYSQL in Laravel frameworks.CO 6: Create applications utilizing frameworks to solve real-time problem. (K6)

Unit-I **(6 Hrs)**
ESSENTIAL PHP: Creating your Development Environment- Mixing HTML and PHP – Command Line PHP – Working with Variables – Creating Constants – Understanding PHP‟s Internal Data Types – Operators and Flow Control.

Unit-II **(6 Hrs)**
STRINGS AND ARRAYS: String Function – Modifying Data in an Array – Deleting Array Elements – Array with Loops – PHP Array Functions – Sorting Array – Splitting Array – Merging Array.

Unit-III **(6 Hrs)**
CREATING FUNCTION: Passing Function – Passing Arrays to Function – Passing by Reference – Using Default Arguments – Passing Variable Numbers of Argument – Returning Data from Function - Nesting Functions.

Unit-IV **(6 Hrs)**
WORKING WITH DATABASE: Creating a MYSQL Database – Creating a New Table – PuttingData into the New Database – Accessing the Database –Update data into the Database– Insert datainto the Database – Delete data from Database– Handling and Avoiding Errors.

Unit-V **(6 Hrs)**
LARAVEL FRAMEWORK OVERVIEW: Introduction - Advantages of Laravel - Features ofLaravel.Laravel -Installation - Application Structure–Configuration – Routing – Controllers – Request - Response - Laravel Forms and HTML Component.

List of Practicals
1. Creating simple webpage using PHP
2. Use of conditional statements in PHP
3. Use of looping statements in PHP
4. Usage of array functions
5. File manipulation using PHP
6. String Functions and Arrays
7. Functions using PHP
8. Create a session and cookies
9. Form Validation using PHP
10. Database connectivity in PHP with MySQL
11. Insertion, Updation and Deletion of rows in MYSQL tables
12. JavaScript validation

Text Books:

1. Jon Duckett, "Beginning HTML, XHTML, CSS, and JavaScript", Wiley Publishing, Indiana, 2010.
2. James Lee, Brent Lee “Open-Source Development with LAMP Using Linux, Apache, MySQL, Perl and PHP”, Pearson Education, 2009.
3. Matt Stauffer, “LARAVEL Up and Running, A Framework for building modern PHP Apps”,2nd Edition, O'REILLY, 2019.

Reference Books:

- John Dean, "Web Programming with HTML5, CSS and JavaScript" Jones & Bartlet Learning, 2019.
- Json Gerner, Elizabeth Naramore, Morgan Owens and Matt Warden, "Professional LAMP - Using Linux, Apache, My SQL and PHP5 Web development", Wiley Publisher, 2006.
- Jack Vo, Learning Laravel: The Easiest Way, Learninglaravel.net, 2014.

MySQL
(Course Code: 23PDSCC3)

SEMESTER– II	HOURS:30	CREDITS – 3

Objective:

Understand basic concepts of how a database stores information via tables.

Outcomes:

At the end of the course the students will be able to

CO1: Remember the basic concepts of database in MYSQL. (K1)
CO2: Understand the basic concepts of ER model and DDL commands (K2)
CO3: Apply the queries to manipulate the data in MYSQL. (K3)
CO4: Analyze and Understand and the concept to implement the join and set operation In MYSQL. (K4)
CO5: Create a PL/SQL program to evaluate the SQL queries in MYSQL. (K6)
CO6: Evaluate a PL/SQL program using Control Statements. (K6)

Unit I: RELATIONAL DATABASE AND DATA BASE DESIGN (6 Hours)

Purpose of Database systems-View of Data-Database Languages-Relational databases - Structure of relational databases-Data Base Architecture-Data Base users and Administrators- Overview of design process-E-R Model-E-R Diagram.

Unit II: DATA DEFINTION LANGUAGE (6 Hours)

Data Types-Constraints-Creating an Oracle Table-Displaying Table Information-Altering Existing Table-Dropping Table-Renaming Table-Truncating Table-Spooling.

Unit III: DATA MANIPULATION AND RETRIEVAL (6 Hours)

Data manipulation Language - Adding a New Row/Record-Updating Existing Rows/Records-Deleting Existing Rows/Records -Retrieving Data from a table-Arithmetic Operators-Group by Having Clause

Unit IV: JOINS AND SET OPERATORS (6 Hours)

Join-Cartesian Product-Equi Join-Outer Join-Set Operators-Nested Queries-Subquery- Queries-Top-N Analysis-Correlated Subquery- Controlling access privileges.

Unit V: NORMALIZATION (6 Hours)

First Normal Form-Second Normal Form-Third Normal Form-BCNF -Fifth Normal form.

List of Exercises:

1. Creating Tables - Creating, Viewing, Altering, Modifying, Dropping, Renaming, Truncating.
2. Manipulating Tables- Insert, Update, Delete.
3. Working with Decode and Case.
4. Retrieving Rows with Character Function
5. Retrieving Rows with Number and Date Functions.
6. Retrieving Rows with Group Function and HAVING.
7. Retrieving Rows with Sub Queries.
8. Queries Using Joins
9. Normalization.

Text Books:

1.Daniel Nichter,Efficient MySQL Performance,Best Practices and Techniques,FirstEdition, O'REILY publication, 2021.

Reference Books:

1. Charles Bell,MatsKindaahl,My SQL high availability:Tools for Building Robust DataCenters, SecondEdition,O'Reily, 2014.
2. Charles A Bell, Expert MySQL, 2nd Edition, Apress, 2013.

Web Resources

1. http://www.w3schools.com
2. http://www.tuorialspoint.com
3. http://www.javapoint.com

NoSQL
(23PDSCC4)

SEMESTER– II	HOURS:30	CREDITS – 3

Objective:

To improve the knowledge on the basic concepts of Relation database management system and learn about SQL language and its commands.

Outcomes:

At the end of the course the students will be able to:

CO1: Compare the concepts of RDBMS and NoSQL. (K4)
CO2: Apply the database modeling techniques and SQL commands. (K3)
CO3: Apply the Queries to manipulate the data.(K3)
CO4: Analyze the concepts to implement set operations.(K4)
CO5: Create a program to solve real-time problems using SQL queries. (K6)
CO6: Evaluate programs and queries using Control Statements.(K6)

Unit I: (12 Hours)

Relational Database and Data Base Design: View of Data-Database Languages-Relational databases - Structure -Data Base Architecture-Data Base users and Administrators- Design Process-E-R Model.**Key Value Data Stores:** NoSQL Key/Value databases using MongoDB, Document Databases, Document oriented Database Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases.

Unit II: (12 Hours)

Data Definition Language: Types of NoSQL Database - Persistent Data – Concurrency – Integration - Impedance Mismatch - Application and Integration Databases - Attack of the Clusters - Emergence of NoSQL - stores - Mongo DB - Cassandra - HBASE - Neo4j use and deployment - Application.

Unit III: (12 Hours)

Data Manipulation and Retrieval: Event Logging - Content Management Systems - Blogging Platforms - Web Analytics or Real-Time Analytics - E-Commerce Applications - Spanning Different Operations - Queries against Varying Aggregate Structure.

Unit IV: (12 Hours)

Joins and Set Operators: Document Oriented Database: Column- oriented NoSQL databases using Apache HBASE, Column-oriented NoSQL databases using Apache Cassandra, Architecture of HBASE, Column-Family Data Store Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Event Logging, Content Management Systems, Blogging Platforms, Counters, Expiring Usage.

Unit V: (12 Hours)

PL/SQL: Data Modeling with Graph: Comparison of Relational and Graph Modeling, Property Graph Model Graph Analytics: Link analysis algorithm- Web as a graph, Page Rank Markov chain, page rank computation, Topic specific page rank Page Ranking Computation techniques iterative processing, Random walk distribution Querying Graphs.

Text Books:

1. Sadalage, P., Fowler, NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Wiley Publications, 1st Edition,2022

Reference Books

1. Guy Harrison, Next Generation Database: NoSQL and Big Data, Apress.
2. Alexix Leon and Mathews Leon,Fundamentals of Data Base Management Systems,Vijay Nicole Imprints,2010.

Web Resources

1. http://www.w3schools.com
2. http://www.tuorialspoint.com
3. http://www.javapoint.com

BUSINESS ANALYTICS LAB
(Course Code: 23PDSCC5)

SEMESTER– III	HOURS –30	CREDITS – 3

Objective:

To provide students with a comprehensive understanding of the field of data science and business analytics, including data management and visualization, data science project lifecycle, machine learning fundamentals, and practical applications in different business domains.

Outcomes:

At the end of the course the students will be able to

CO1: Describe the basics of business analysis and Data Science Knowledge (K2)
CO2: Associate the data management and handling with Data Science Project Life Cycle (K2)
CO3: Apply the data mining concept and its techniques Applying (K3)
CO4: Evaluate the machine learning concepts used for effective analytics (K5)
CO5: Apply the knowledge of business analysis in different domain Applying (K4)
CO6: Design models useful for analysis in the field of business. (K6)

Unit I: (6 Hrs.)

Introduction: Historical Overview - Data Scientist vs. Data Engineer vs. Business Analyst, Career in Business Analytics - data science - Applications for data science, Data Scientists Roles, and Responsibility.

Unit II: (6 Hrs.)

Data: Data Collection, Data Management, Big Data Management, Organization/sources of data, Importance of data quality, Dealing with missing or incomplete data, Data Visualization, Data Classification. Data Science Project Life Cycle: Business Requirement, Data Acquisition, Data Preparation, Hypothesis and Modeling, Evaluation and Interpretation, Deployment, Operations, Optimization.

Unit III: (6 Hrs.)

Data Mining: The origins of Data Mining, Data Mining Tasks, OLAP and Multidimensional data analysis, Basic concept of Association Analysis and Cluster Analysis.

Unit IV: (6 Hrs.)

Machine Learning: History and Evolution, AI Evolution, Statistics Vs Data Mining Vs, Data Analytics Vs, Data Science, Supervised Learning, Unsupervised Learning, Reinforcement Learning, Frameworks for building Machine Learning Systems.

Unit V: (6 Hrs.)

Application of Business Analysis: Retail Analytics, Marketing Analytics, Financial Analytics, Healthcare Analytics, Supply Chain Analytics.

Lab Activities

1. Exploratory Data Analysis
2. Create visualizations
3. Data Cleaning and Preprocessing
4. Regression Analysis
5. Hypothesis Testing
6. Clustering Analysis
7. Classification Algorithms

8. Time Series Forecasting
9. Business Intelligence Dashboard
10. Text Analytics
11. A/B Testing Experiment
12. Data Storytelling Presentation

Text Books:

1. Bhimasankaram Pochiraju, Sridhar Seshadri, Essentials of Business Analytics: An Introduction to the methodology and its application, Springer, 2019.
2. Andreas C. Müller, Sarah Guido, Introduction to Machine Learning with Python: A Guide for Data Scientists, 1st Edition, O'Reilly, 2016.
3. Laura Igual Santi Seguí, Introduction to Data Science, Springer, 2017.

Reference Books:

- Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Education India, 2021.
- Ger Koole, An Introduction to Business Analytics, Lulu.com, 2019.

HEALTH CARE ANALYTICS
(Course Code: 23PDSCC6)

SEMESTER– III	HOURS –30	CREDITS – 3

Objective:

To understand the use of analytics techniques in health care and explore the concepts such as the nature of heath data, scaling, and evaluate a model.

Outcomes:

At the end of the course the students will be able to:

CO1: Describe the fundamental concepts and techniques of data analytics, demonstrating knowledge and comprehension. (K2)

CO2: Associate algorithms to real-world datasets, showcasing application and analysis skills (K2).

CO3: Examine the performance and effectiveness of models using appropriate metrics, engaging in critical thinking and evaluation. (K3)

CO4: Analyse algorithms to identify meaningful patterns and relationships in large datasets, demonstrating synthesis abilities. (K4)

CO5: Assess techniques to group similar data points and interpret the results, demonstrating analysis and synthesis skills. (K5)

CO6: Design models or algorithms useful for predictive analysis in health care industry. (K6)

Unit I: **(6 hrs)**

Introduction: Data Analytics - Introducing R software - Exploring Data - Classification: - Basic Concepts- Decision Trees - Model Evaluation.

Unit II: **(6 hrs)**

Classification: Rule Based - Nearest Neighbors - Support Vector Machines - Bayesian Networks - Neural Networks.

Unit III: **(6 hrs)**

Association Analysis: Basic Concepts and Algorithms - Cluster Analysis: Basic Concepts and Algorithms.

Unit IV: **(6 hrs)**

Regression Methods: Multiple Linear Regression - Logistic Regression - Optimization - Linear Programming - Integer Programming - Text Mining.

Unit V: **(6 hrs)**

Fundamentals of Healthcare Analytics: Improve Decision Making - Analytics, Quality, and Performance - Applications – Components - Working with Data - Raw Material - Analyzing Data.

Lab Exercises

1. Exploratory Data Analysis on Health Data Sets
2. Create visualizations using Health Data Sets
3. Data Cleaning and Preprocessing in Health Data Sets collected from different sources
4. Regression Analysis on Health Data Sets
5. Clustering Analysis in Health Data Sets
6. Working with Classification Algorithms on Health Data Sets
7. Dashboard for Monitoring Health Data
8. Text Analytics for patient monitoring

Text Books:

1. P.N. Tan, M. Steinbach, V. Kumar., Introduction to Data Mining, Pearson Addison Wesley, 2006.
   Unit I: Ch.1, Ch.2, Ch. 3, Ch. 4, **Unit II:** Ch. 5, **Unit III:** Ch.6, **Unit IV:** Ch.8
2. Trevor L Strome, Healthcare Analytics for Quality and Performance Improvement, Wiley, 2013.
   Unit V: Ch. 2, Ch. 6

Reference Books:

1. Elements of Statistical Learning, by T. Hastie, R. Tibshirani, J. Friedman, Springer Verlag, NY, 2001.
2. Machine Learning, by T. Mitchell, McGraw Hill, NY, 1997.
3. Reddy, Chandan K., and Charu C. Aggarwal, eds. Healthcare data analytics. Vol. 36. CRC Press, 2015.
4. Strickland, Jeffrey. Predictive Analytics Using R. Lulu.com, 2015.

Web References

- https://www.datacamp.com/blog/all-about-r
- https://cran.r-project.org/manuals.html

MULTIMEDIA I
(Course Code: 23PDSCC7)

SEMESTER– IV	HOURS –30	CREDITS – 3

Objective:

To be able to use Image Manipulation Programs for digital image editing and manipulation, enabling students to create visually compelling and professional-quality graphics.

Outcomes:

At the end of the course the students will be able to

CO1: Classify the GIMP's features, tools, and interface, and utilize them effectively for image editing and manipulation. (K4)

CO2: Apply a wide range of editing techniques to enhance and retouch digital images, including color correction, retouching, and applying filters and effects. (K2)

CO3: Create and design graphics using GIMP, utilizing drawing tools, shapes, and text, and apply formatting and effects to achieve desired visual results. (K3)

CO4: Evaluate the components such as layers, masks, and blending modes to create complex image compositions and manipulations. (K5)

CO5: Illustrate modifying images for different platforms and outputs, including web and print, and export images in various formats while considering file size and quality. (K4)

CO6: Design visually appealing and engaging graphics. (K6)

Unit I: **(6 hrs)**

Introduction to GIMP: Overview - GIMP interface - layers and image manipulation tools - Selection and transformation techniques - Basic editing and navigation shortcuts.

Unit II: **(6 hrs)**

Image Editing and Enhancement: Editing in GIMP - Color correction and adjustment tools - Retouching and removing blemishes - Applying filters and effects - Masks and blending modes.

Unit III: **(6 hrs)**

Creating Graphics and Text: Graphic design in GIMP - Creating shapes – Lines - Curves - Adding and formatting text - Brushes and Patterns - Logos and Simple Illustrations.

Unit IV: **(6 hrs)**

Advanced Techniques: Working with advanced selection tools - Image composition and manipulation - Creating photo montages and collages - Utilizing paths and bezier curves - Implementing perspective and distortion effects.

Unit V: **(6 hrs)**

Exporting and Output: Optimizing images for web and print - Exporting images in different formats - Introduction to batch processing - Printing and print preparation - Creating GIF animations.

Lab Activities

1. Designing an Employee or Student ID card
2. Designing a seasonal greeting
3. Design a photograph applying Filter effects
4. Design an invitation for a conference
5. Design a brochure or poster for a technical symposium

6. Designing a Flexible banner for your college
7. Create your own wallpaper for your Desktop/Laptop/Mobile
8. Design a Web banner for a website

Text Books:

1. Akkana Peck, "Beginning GIMP: From Novice to Professional", 2nd Edition, Apress, 2005.
2. Jason van Gumster, Robert Shimonski, "GIMP Bible", Wiley, 2010. ISBN: 978-0-470-52397-1

Reference Books:

- Klaus Goelker, "GIMP 2.8 for Photographers: Image Editing with Open-Source Software", Rocky Nook, 2013.
- Carey Bunks, "Grokking the GIMP", Sams, 2000.

MULTIMEDIA II

(Other Department Students)

(Course Code: 23PDSCC8)

SEMESTER– IV	HOURS –30	CREDITS – 3

Objective:

To equip students with the skills and knowledge to create dynamic and interactive web animations using Synfig and Aptana.

Outcomes

At the end of the course the students will be able to:

CO1: Apply the principles and techniques of animation using Synfig. (K2)

CO2: Experiment using keyframes, interpolation, masking, and rigging techniques in Synfig. (K4)

CO3: Develop proficiency in using Aptana for web development, including creating and styling web pages using HTML, CSS, and JavaScript. (K3)

CO4: Apply animation to contents in web pages and add interactivity. (K4)

CO5: Design responsive web layouts and optimize web animations for different devices and screen sizes. (K5)

CO6: Deploy web animations to hosting servers, optimize performance, and test across multiple browsers. (K6)

Unit I: **(6 hrs)**

Animation with Synfig: Animation principles and techniques - Synfig software interface and features - Creating basic animations using keyframes and interpolation – layers – shapes - transformations.

Unit II: **(6 hrs)**

Advanced Techniques: masking and morphing - Utilizing bones and rigging for character animation - Incorporating sound and video elements into animated projects - Exporting animations for various platforms and formats.

Unit III: **(6 hrs)**

Web Development Fundamentals: Web development languages - Understanding the structure and components of a website - Utilizing Aptana IDE for web development projects - Creating and styling web pages using HTML and CSS.

Unit IV: **(6 hrs)**

Interactive Web Animation: Incorporating animated elements into web pages using Synfig and Aptana - Implementing interactivity and user engagement through JavaScript - Creating responsive web designs and layouts - Optimizing web animations for different devices and screen sizes.

Unit V: **(6 hrs)**

Web Animation Deployment and Optimization: Deploying animations to servers - Performance optimization techniques for web animations - Testing and debugging web animations in different browsers - Collaborative tools.

Lab Activities

Synfig

1. Create an animation for bouncing a ball
2. Create brushed outlines for an image
3. Build a magnifying glass
4. Develop a slide show of photos with transitions

Aptana

1. Developing a simple webpage with images and links
2. Develop a webpage displaying the timetable of the Department
3. Design an application form for Student Admission
4. Create your own web blog for college events

Text Books:

1. Ankit Ahuja, "Synfig Studio: 2D Animation", AMC College, Information and Communication Technology (ICT).
2. Terry Felke-Morris, "Web Development and Design Foundations with HTML5", Pearson, 2021.

Reference Books:

- Richard Williams, "The Animator's Survival Kit", Faber & Faber, 2009.
- "HTML and CSS: Design and Build Websites" by Jon Duckett
- "JavaScript and JQuery: Interactive Front-End Web Development" by Jon Duckett
- "Learning Web Design: A Beginner's Guide to HTML, CSS, JavaScript, and Web Graphics" by Jennifer Niederst Robbins
- "CSS Secrets: Better Solutions to Everyday Web Design Problems" by Lea Verou

Extra Credit Course

MICROSOFT EXCEL

(Other Department Students)

(Course Code: 23PDSEC1)

SEMESTER– I	HOURS –30	CREDITS – 3

Objective:

To provide students with advanced knowledge and skills on MS Excel to effectively work with basic operations, functions, and design insightful visualizations.

Outcomes:

At the end of the course the students must be able to:

CO1: Experiment the Basic functions in Excel. (K2)
CO2: Create a simple worksheet in Excel(K6)
CO3: Formulate the Basic formulas and functions in Excel. (K6)
CO4: Apply various functions and formulas for different data. (K3)
CO5: Generate a simple chart in Excel. (K6)
CO6: Compare the data using Mathematical functions. (K4)

UNIT I: (6 hrs)

Introduction to Excel: Understanding Workbooks- Navigating-File- Typing Text Numbers into Worksheet- Typing Simple Formulas in A Worksheet- Inserting and Deleting Worksheets.

UNIT II: (6 hrs)

Changing Worksheet Tab Colours- Freezing Rows and Columns- Selecting Ranges, Rowsand Columns- Changing Fonts- Understanding Borders-Applying A Border to A Range-Wrapping and Merging Text.

UNIT-III (6 hrs)

Building Basic Formulas - Formula building methods- Editing formulas- Building formulas with computational operators

UNIT IV (6 hrs)

Building Basic Formulas-Creating Formulas.: Using Formulas, Formula Functions – Sum, Average, if, Count, max, min, Proper, Upper, Lower. Using AutoSum- Logical Functions-Understanding Error Values.

UNIT V (6 hrs)

Trim Spreadsheet Charts: Creating Charts, Different types of charts, Formatting Chart Objects, Changing the Chart Type, Showing and Hiding the Legend, Showing and Hiding the Data Table

Text Books:

1. Greg Harvey, Microsoft Excel 2019 All in One: John Wiley & Sons, Inc.2019 edition
2. Joseph Schmuller, Statistics Anaysis with Excel, John Wiley & Sons, 3rd Edition.
3. Stephen L.Nelson, Microsoft Data Analysis, John Wiley & Sons, 3rd Edition.

Web References:

1. https://www.analyticsvidhya.com/
2. https://www.simplilearn.com/tutorials/excel-tutorial/data-analysis-excel
3. https://www.coursera.org/learn/excel-basics-data-analysis-ibm

MICROSOFT OFFICE TOOLS
(Other Department Students)
(Course Code: 23PDSEC2)

SEMESTER– II	HOURS –30	CREDITS – 3

Objective:

To be able to effectively use word processing and presentation software to create visually appealing documents and presentations, incorporating various content types, applying formatting techniques, and utilizing advanced features such as collaboration, animation, and object manipulation.

Outcomes

At the end of the course the students will be able to:

CO1: demonstrate proficiency in document formatting, manipulation, and efficient text selection techniques. (K3)

CO2: **A**pply advanced features like sorting, collaboration, compatibility checking, and charts for representing the data using Excel. (K3)

CO3: Comprehensive understanding of the PowerPoint interface, enabling them to create visually appealing presentations with well-designed slides and utilize object formatting and grouping techniques. (K3)

CO4: Experiment proficiency in incorporating various content types, such as tables, charts, graphics, and media clips, into PowerPoint presentations, and effectively apply slide design principles to create impactful and visually engaging slides. (K4)

CO5: Analyze and gain expertise in animation, transitions, including shapes, and diagrams, and utilize presenter tools to deliver dynamic and engaging presentations. (K5)

CO6: Create documents including journal articles or books and design animations and annotated presentations with info graphs (K6)

Unit I: **(6 hrs)**

Getting Familiar and Document Navigation: Interface Overview - Common Buttons - Tabs - Ribbons - Flashing Line – Slider - Keyboard Navigation - Browse Objects Tools - Selecting Text Options. Manipulating Text and Formatting: Page Layout - Cut, Copy, Paste, Move, Find and Replace - Changing Fonts - Typical Text Format Options - Bullet and Numbered Lists - Other Formatting Tools – Line Spacing. Themes and Templates: Applying Themes - Opening - Creating - Working with Existing Content – Options - Inserting Fields – Inserting Images.

Unit II: **(6 hrs)**

Organizing Content: Creating Lists - Creating and Formatting Tables - convert text to tables - Table Borders and Shading - Creating Columns - Creating Tabular Lists - Creating Charts - Character Spacing - Inserting Shapes - Inserting Images - Word Art - Text Boxes - Sorting Text - Smart Art - Document Views: Split Screen View - Tracking Changes - Collaboration options - Combining and Approving Changes - Using Comments - Compatibility Checker - Previous Word Versions Compatibility - Spell and grammar Check - Saving a Document - Printing and Preview Options.

Unit III: **(6 hrs)**

Getting acquainted with PowerPoint: Power Point interface - Outline and Slides pane – Ribbon - Quick Access toolbar - notes pane - Status Bar. Objects and Formatting: Objects in PowerPoint - Copying and moving objects - Format painter - Fonts and effects - Inserting new slides - Slide layout – Selecting multiple objects - Grouping objects.

Unit IV: **(6 hrs)**

Six content types: Table – Chart – Graphic – Picture - Clip art - Media clip - Slide Design - Templates and Slide Masters - themes, masters and templates - Layouts and footers - Principles - Making slides clearer - slides with real impact. Making an Impact: Use images - insert and edit - Cropping - resizing - manipulating images - Screenshots and videos - Removing picture backgrounds - Applying artistic effects.

Unit V: **(6 hrs)**

Animation: Adding and controlling transition effects – Detailed animation and transition options – animating – text boxes – charts – diagrams. Presenter tools and features: Custom slide shows – Annotating slide – Navigating a presentation – Self-running presentations.

Text Books:

1. Alexander Mamishev, Murray Sargent, "Creating Research and Scientific Documents Using Microsoft Word", Microsoft Press, 2014.

Unit I, Unit II

2. Joan Lambert, "Microsoft PowerPoint step by step 2016", Microsoft Press, 2016.

Unit III, Unit IV, Unit V

Reference Books:

- Ann Shaffer and Katherine T. Pinard, "New Perspectives Microsoft Office 365 & PowerPoint 2016: Intermediate", Course Technology, 2016.
- Joan Lambert, "Microsoft PowerPoint step by step 2019", Microsoft Press, 2019.

WEB SCRAPING USING PYTHON
(Data Science Students)
(Course Code: 23PDSEC3)

SEMESTER– III	HOURS –30	CREDITS – 3

Objective:

To provide students with advanced skills in web scraping using Python, enabling them to scrape complex websites, handle dynamic content, analyze data, and adhere to ethical considerations.

Outcomes:

At the end of the course the students will be able to:

CO1: Describe the techniques to scrape websites with complex HTML structures (K2)

CO2: Apply the knowledge on retrieving data from websites built with JavaScript frameworks, such as React, Angular, and Vue.js. (K3)

CO3: Analyze the advanced data extraction techniques, including handling authentication, overcoming CAPTCHA challenges, and extracting data from APIs and REST ful services, showcasing analytical and problem-solving skills (K3).

CO4: Experiment preprocess, and analyze web-scraped data, employing exploratory data analysis techniques and performing feature engineering (K4).

CO5: Create visualize and communicate insights from web-scraped data effectively, employing advanced data visualization techniques and storytelling principles (K6).

CO6: Evaluate the ethical considerations and adhere to best practices in web scraping (K5).

Unit I: (6 hrs)

Introduction to Advanced Web Scraping: Review of basic web scraping concepts and tools - Understanding the importance of advanced web scraping techniques - Handling complex HTML structures and nested data - XPath and CSS selectors for advanced data extraction - Handling dynamic content and AJAX calls.

Unit II: (6 hrs)

Scraping with JavaScript Frameworks: Overview of frameworks (React, Angular, Vue.js) - Approaches to scraping websites built with JavaScript frameworks - Using headless browsers (Selenium, Puppeteer) for dynamic rendering - Extracting data from single-page applications (SPAs) - Dealing with virtual DOM manipulation.

Unit III: (6 hrs)

Advanced Data Extraction Techniques: Handling authentication and login forms - Scraping websites behind CAPTCHA challenges - Extracting data from APIs and RESTful services - Parsing and extracting data from XML and JSON responses - Working with complex data structures (nested JSON, arrays).

Unit IV: (6 hrs)

Data Processing and Analysis: Cleaning and preprocessing scraped data at scale - Exploratory data analysis techniques for web scraping data - Feature engineering and data transformation - Advanced data visualization and storytelling with scraped data - Introduction to machine learning for web scraping data analysis.

Unit V: (6 hrs)

Ethical Considerations and Best Practices: Ethical considerations in web scraping (terms of service, scraping etiquette) - Handling anti-scraping measures and IP blocking - Scaling and optimizing web scraping tasks - Deploying web scrapers in production environments - Best practices for maintaining and updating web scrapers.

Text Book:

1. Ryan Mitchell, "Web Scraping with Python: Collecting More Data from the Modern Web ", 2nd Edition, O'Reilly, 2018.

Reference Books:

- Andy Vickler, Python: Python Automation Techniques and Web Scraping, Kindle Edition, 2021.
- Brian Jones, David Beazley, Python Cookbook: Recipes for Mastering Python 3, Third Edition, O'Reilly, 2022.

FINANCIAL ANALYTICS USING EXCEL
(Data Science Students)
(Course Code: 23PDSEC4)

SEMESTER– IV	HOURS –30	CREDITS – 3

Objective:

To develop expertise in Excel-based financial modeling, encompassing data preparation, statistical analysis, investment evaluation, risk assessment, scenario analysis, and impactful report creation.

Outcomes:

At the end of the course the students will be able to:

CO1: Create and manipulate financial models using Excel. (K6)
CO2: Describe how to import, clean, prepare, and transform financial data for analysis. (K3)
CO3: Apply statistical analysis on financial data using Excel. (K4)
CO4: Evaluate investment opportunities using Excel, including techniques such as NPV, IRR, and sensitivity analysis. (K5)
CO5: Demonstrate the use of Excel for risk assessment and scenario analysis. (K4)
CO6: Create effective financial reports and visualizations for conveying insights. (K6)

Unit I: (6 hrs)

Introduction to Financial Analytics and Excel Basics: Introduction to financial analytics and its importance in decision-making-Excel basics: navigating the interface, cell referencing, basic formulas, formatting-Working with data: sorting, filtering, and data validation-Introduction to functions: SUM, AVERAGE, COUNT, MIN, MAX.

Unit II: (6 hrs)

Time Value of Money and Investment Analysis: Understanding time value of money (TVM) concepts-Calculating present and future values, interest rates, and periodic payments-Investment analysis using Excel: NPV, IRR, ROI-Using Excel's built-in financial functions: PV, FV, PMT, RATE.

Unit III: (6 hrs)

Financial Statement Analysis: Introduction to financial statements: balance sheet, income statement, cash flow statement. **Ratio analysis:** liquidity, profitability, solvency, efficiency-Constructing financial statements in Excel-Using Excel functions for financial ratios: ROA, ROE, current ratio.

Unit IV: (6 hrs)

Forecasting and Budgeting: Time series analysis and forecasting techniques-Building budgets and financial projections in Excel-Moving averages, exponential smoothing, trend analysis- Valuation and Sensitivity Analysis-Valuation methods: DCF (Discounted Cash Flow), comparables, precedent transactions-Performing sensitivity analysis to assess the impact of changing assumptions-Monte Carlo simulation for risk assessment-Building valuation models in Excel.

Unit V: (6 hrs)

Advanced Financial Modeling: Building complex financial models: mergers and acquisitions, project finance-Data consolidation and analysis using PivotTables-**Advanced Excel functions:** INDEX, MATCH, OFFSET, array formulas-Creating interactive dashboards for financial reporting.

Text Books:

1. Danielle Stein Fairhurst "Financial Modeling in Excel for Dummies", 2nd, Kindle Edition, For Dummies, 2021.
2. Paul Mc Fedries "Excel Data Analysis for Dummies", For Dummies, 2022. ISBN: 978-1119844426.

Reference Books:

1. Michael Taillard, "Corporate Finance for Dummies", 2nd Edition, Wiley, 2023.
2. Stephen H. Penman, "Financial Statement Analysis and Security Valuation", 4th Edition, Mc Graw Hill Education, 2009.
3. Krishna G. Palepu, Paul M. Healy, "Business Analysis and Valuation: Using Financial Statements", Cengage India Pvt. Ltd., 5th Editon, 2015.