M.TECH DataScience

Scheme of Instruction and Syllabus

w.e.f 2023– 24

Department of Information Technology

VELAGAPUDI RAMAKRISHNA SIDDHARTHA ENGINEERING COLLEGE

(An Autonomous Institution affiliated to Jawaharlal Nehru Technological University Kakinada, Kakinada.
ISO 21001:2018 Certified)
(Sponsored by Siddhartha Academy of General & Technical Education)

Kanuru, Vijayawada-520007, A.P. India

Curriculum Structure for M.TECH DataScience   w.e.f 2023– 24

SEMESTER I   **Contact Hours: 26**

S.No	Course Type	Course Code	Title/Type of the Course	L	T	P	C
1	Programme Core - I	23ITDS1001	Mathematical Foundations For Data Science	3	0	0	3
2	Programme Core - II	23ITDS1002	Advanced Data Structures and Algorithms	3	0	0	3
3	Programme Core - III	23ITDS1003	Machine Learning	2	0	2	3
4	Programme Elective - I	23ITDS1014	**A.** Statistics with R **B.** Advanced Java Programming **C.** Data analysis with Python	3	0	0	3
5	Programme Elective - II	23ITDS1015	A. Data Science for Decision Making B. Cloud Data Engineering C. Cyber security and Forensics	3	0	0	3
6	Mandatory Learning Course	23MTMC1026	Research Methodology and IPR	2	0	0	0
7	Laboratory - I	23ITDS1051	Advanced Data Structures and Algorithms Lab	0	0	3	1.5
8	Laboratory - II	23ITDS1052	Program Elective -1 Lab	0	0	3	1.5
9	Project	23ITDS1063	Capstone Project 1 (PC-III)	0	0	2	1
			Total	**16**	**0**	**10**	**19**

SEMESTER II   **Contact Hours: 28**

S.No	Course Type	Course Code	Title/Type of the Course	L	T	P	C
1	Programme Core–IV	23ITDS2001	Big Data frameworks for Data Science	3	0	0	3
2	Programme Core – V	23ITDS2002	Deep Learning	2	0	2	3
3	Programme Core – VI	23ITDS2003	Data Visualization and Interpretation	3	0	0	3
4	Programme Elective – III	23ITDS2014	A. Business Analytics and Modelling B. Image and Video Analytics C. Natural Language Processing	3	0	0	3
5	Programme Elective – IV	23ITDS2015	A. Web mining and Social Network Analysis B. Optimization Techniques for Data Analysis C. Information Retrieval Systems	3	0	0	3
6	Audit Course	23MTAC2036	Technical Report Writing	2	0	0	-
7	Term Paper	23ITDS2063	Term Paper[2]	2	0	0	1
8	Laboratory - I	23ITDS2051	Big Data Lab	0	0	3	1.5
9	Laboratory - II	23ITDS2052	Data Visualization Lab	0	0	3	1.5
10	Project	23ITDS2064	Capstone Project 2 (PC-V)	0	0	2	1
			Total	**18**	**0**	**10**	**20**

Semester III   **Contact Hours:23**

S.No	Course Type	Course Code	Title/Type of the Course	L	T	P	C
1	Programme Elective - V	23ITDS3011	MOOCS Course	3	0	0	3
2	Project (Part-A)	23ITDS3061	Dissertation/ Industrial Project	0	0	20	10
3	Internship	23ITDS3052	Internship/Summer Training in Research Organizations/ Institutions of Higher Learning (After II Sem)	0	0	0	2
			Total	**3**	**0**	**20**	**15**

Semester IV   **Contact Hours:32**

S.No	Course Type	Course Code	Title/Type of the Course	L	T	P	C
1	Project (Part-B)	23ITDS4061	Dissertation/ Industrial Project	0	0	32	16
			Total	**0**	**0**	**32**	**16**

Semester	Credits
1	19
2	20
3	15
4	16
Total	**70**

L – Lecture, T – Tutorial, P – Practical, C – Credits

Note:
1. Student has to carryout a project applying the knowledge and hands on technical skills they have gained through course work and lab sessions in **Semester-I** under **Capstone Project 1**
2. Student should carryout literature survey of the selected problem and present it in a Seminar for the yearlongProject Work under Term Paper.
3. Student has to carry out a project applying the knowledge and hands on technical skills they have gained through course work and lab sessions in **Semester-II** under **Capstone Project 2**
4. At least one theory course in I&IIsemesters can be made as integratedcourse(Theory coupled with Laboratory).
5. Maximum of three theory courses (40% of courses) can be offered as self-learning courses in each of the First and Second semesters.

SEMESTER I

23ITDS1001 - MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE

Course Category:	Programme Core-I	Credits:	3
Course Type:	Theory	Lecture-Tutorial-Practice:	3-0-0
Prerequisites:	Basic Mathematics	Continuous Evaluation:	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	Upon successful completion of the course, the student will be able to:	
	CO1	Understand basic mathematical concepts like calculus and linear algebra
	CO2	Derive the probability mass and density functions of transformation of random variables
	CO3	Apply the mathematical and probabilistic foundations of statistical inference in computing
	CO4	Interpret the results of Regression and Correlation Analysis, for forecasting , perform analysis of variance

Contribution of Course Outcomes towards achievement of Program Outcomes(1-Low, 2-Medium, 3- High)		PO 1	PO 2	PO 3	PO 4	PO 5		
	CO1	1			3			
	CO2				3			
	CO3	2			2	1		
	CO4	1		1	1	2		

Course Content	**UNIT I:** **Derivatives:** Derivatives and rates of change, The derivative as a function, Differentiation formulas, Derivatives of trigonometric functions, The chain rule, Implicit differentiation. Applications of Differentiation, Maximum and minimum values, The mean value theorem, How derivatives affect the shape of a graph, summary of curve sketching, Graphing with calculus and calculators, Integrals,The definite integral, The fundamental theorem of calculus, The substitution rule. **Linear Algebra:** Introduction to Vectors, solving linear equations **Eigen values and Eigen vectors:** Diagonalzing a matrix, Systems of Differential Equations.
	UNIT II: **Probability and statistics:** Introduction- Probability models, sample space, events, algebra of events, graphical methods of representing events, probability axioms, conditional probability, independence of events. **Discrete Random Variables**:The Probability Mass Function, Special Discrete Distributions, The Probability Generating Function, Independent Random Variables **Continuous Random Variables**: Introduction, The Exponential Distribution, The Reliability and Failure Rate, Some Important

	Distributions, Functions of a Random Variable, Jointly Distributed Random Variables, Order Statistics.
	UNIT III: **Expectation:** Introduction, Moments, Expectation Based on Multiple Random Variables, Transform Methods, Moments and Transforms of Some Distributions, Computation of Mean Time to Failure. **Stochastic Process:** Classification of Stochastic Processes, The Bernoulli Process, The Poisson Process
	UNIT IV: **Statistical Inference:** Introduction, parameter estimation, hypothesis testing **Regression and Analysis of variance:** Introduction, Least-squares Curve Fitting, The Coefficients of Determination, Correlation Analysis, Simple Nonlinear Regression, Higher-dimensional Least-squares Fit, Analysis of Variance.
Text books and Reference books	**Text Book(s):** [1] Calculus, 7th Edition by James Stewart 2015 [2] Gilbert Strang, Linear Algebra and its applications, Wellesley-Cambridge Press, Fifth Edition, 2016 [3] Kishor S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, John Wiley & Sons, 2016 **Reference Book(s):** [1] M. Mitzenmacher and E. Upfal, .Probability and Computing: Randomized Algorithms andProbabilistic Analysis, Cambridge, 2005 [2] John Vince, Foundation Mathematics for Computer Science, Springer, 2015
E-resources and other digital material	[1] Maggie Myers, Robert van de Geijn, (24,06,2019). Linear Algebra - Foundations to Frontiers, UTAustinX, https://www.edx.org/ course/linear-algebra-foundations-to-frontiers-0 [2] Statistics And Probability Tutorial | Statistics And Probability for Data Science | Edureka, https://www.youtube.com/watch?v=XcLO4f1i4Yo [3] Dr Nic's Maths and Stats, Understanding Statistical Inference - statistics help, https://www.youtube.com/watch?v=tFRXsngz4UQ

23ITDS1002- ADVANCED DATA STRUCTURES AND ALGORITHMS

Course Category:	Program Core-II	Credits:		3
Course Type:	Theory	Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	Data Structures	Continuous Evaluation:		40
		Semester end Evaluation:		60
		Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:			
	CO1	Analyze the time and space complexity of the algorithms		
	CO2	Experiment with Tree structures to solve the problems		
	CO3	Develop algorithms using Graph structure to solve real-life problems		
	CO4	Apply suitable data structure and design strategy to solve computing problems		

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	3	3		3	
	CO2		3		3	3
	CO3		3		3	3
	CO4		3		3	3

Course Content	**UNIT I: Algorithm Analysis** Mathematical Background, Model, What to Analyze, Running time calculations-A Simple Example, General Rules, Solutions for the Maximum Subsequence Sum Problem, Logarithms in the Running Time, Limitations of Worst-Case Analysis.
	UNIT II: Trees Preliminaries, Binary Search Trees: Insertion and Deletion, AVL Trees: rotations-insertion and deletion, Splay Trees: Properties-insertion and deletion, Red-Black trees: Properties– Rotations – Insertion – Deletion, B-Trees: Definition, inserting and Deleting a key from a B-Tree.
	UNIT III:Graph Algorithms Representations of Graphs, Breadth-First Search, Depth-First Search, Shortest–Path Algorithms: Dijkstra's Algorithm, Graph with negative edge cost, All-Pairs Shortest Paths, Shortest Path Example, Minimum Spanning Trees: Kruskal's and Prim's algorithm, Applications of Depth-First Search: Finding strong components
	UNIT IV: Algorithm Design Techniques Greedy Algorithms: Optimal Storageon Tapes, Optimal Merge Patterns Dynamic Programming: Multi-stage Graphs, Optimal Binary Search Tree

	Backtracking Algorithms: The 8-QueensProblem, KnapsackProblem Branch-and-Bound: The Method, 0/1 KnapsackProblem, Traveling Salesperson
Text books and Reference books	**Text Book(s):** [1]. Mark Allen Weiss, Data Structures and Algorithm Analysis in C++, 4th Edition, Pearson Publication [2]. Ellis Horowitz, SartajSahni and S. Rajasekharan, Fundamentals of Computer Algorithms, 2nd Edition, Universities Press. **Reference Books:** [1] Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, ―Data Structures and Algorithms‖, Pearson Education,
E-resources and other digital material	[1]. SudarshanIyengar,AssistantProfessor,CSE department, IIT Ropar, Programming, Data Structures and Algorithms [NPTEL], (26, May, 2021) Available: https://nptel.ac.in/noc/courses/noc18/SEM1/noc18-cs25/ [2]. Erik Demaine, professor of Computer Science at the Massachusetts Institute of Technology, Advanced Data Structures [MIT- Open Course Ware], (26, May, 2021) Available: http://ocw.mit.edu/

23ITDS1003-MACHINE LEARNING

Course Category:	Program Core-III		Credits:		3
Course Type:	Integrated Course		Lecture-Tutorial-Practice:		2-0-2
Prerequisites:	---		Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Summarize the fundamental concepts of machine learning			
	CO2	Apply linear, distance based, and decision tree based models for a given scenario			
	CO3	Analyze probabilistic, neural network models			
	CO4	Design a suitable machine learning model for a real world application			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3-High)		PO1	PO2	PO3	PO 4	PO 5
	CO1				1	
	CO2	1			2	1
	CO3	1			2	1
	CO4	2		1	3	2

Course Content	**UNIT I:** **The ingredients of machine learning**: Tasks, Models, Features **Binary classification and related tasks**: Classification, Assessing classification performance, Visualizing classification performance **Beyond binary classification**: Multi-class classification, Regression
	UNIT II: **Decision Tree learning**: Introduction, Decision tree representation, Appropriate problems for decision tree learning, The basic decision tree learning algorithm, Inductive bias in decision tree, Issues in decision tree learning. **Linear models**: The least-squares method, Multivariate linear regression, Support vector machines, Soft margin SVM, Going beyond linearity with kernel methods.
	UNIT III: **Distance Based Models**: Introduction, Nearest Neighbours classification, Distance based clustering, K-Means algorithms, Clustering around medoids, Hierarchical Clustering. **Bayesian Learning**: Introduction, Bayes theorem, Bayes optimal classifier, Naïve Bayes classifier, Bayesian belief networks.
	UNIT IV: **Artificial Neural Networks**: Introduction, appropriate problems for neural network learning, Perceptrons, Multilayer networks and the back propagation **Reinforcement Learning**: Introduction, The Learning task, Q-learning

Text books and Reference books	**Text Book(s):** **[1].** Machine Learning: The art and Science of algorithms that make sense of data, Peter Flach, Cambridge University Press, 2012 **[2].** Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education **Reference (Book)s:** **[1].** AurélienGéron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition **[2].** Stephen Marsland, "Machine Learning – An Algorithmic Perspective", Second Edition, Chapman and Hall/CRC Machine Learning and Pattern Recognition Series, 2014 **[3].** EthemAlpaydın, Introduction to machine learning, second edition, MIT press
E-resources and other digital material	**[1].** Kevin Murphy,"MachineLearning:A Probabilistic Perspective" , MIT Press, 2012, https://www.cs.ubc.ca/~murphyk/MLbook/pml-intro-5nov11.pdf **[2].** Machine Learning by Andrew Ng, Stanford University https://www.coursera.org/learn/machine-learning **[3].** Professor S. Sarkar IIT Kharagpur "Introduction to machine learning" , https://www.youtube.com/playlist?list=PLYihddLFCgYuWNL55Wg8ALkm6u8U7 **[4].** Professor Carl GustafJansson, KTH, Video Course on Machine Learning https://nptel.ac.in/noc/individual_course.php?id=noc19-cs35

23ITDS1014A-STATISTICS WITH R

Course Category:	Programme Elective -I		Credits:		3
Course Type:	Theory		Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	Any programming language and basic Mathematics		Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100

Course Outcomes	Upon successful completion of the course, the student will be able to:	
	CO1	Demonstrate The Semantics, Data Handling And Control Statements In R.
	CO2	Apply Data Manipulation Techniques And Linear, Nonlinear Models On The Given Datasets.
	CO3	Analyze The Relationship Among Data Attributes With Appropriate Techniques.
	CO4	Construct Suitable Plots Using Data Visualizations In R For The Given Application.

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1			1		
	CO2	1		2	3	
	CO3	3		2	3	
	CO4				3	

Course Content	**UNIT I:**
	The R Environment: Command Line Interface, R Studio, Installing R Packages.
	Basics Of R: Basic Math, Variable, Data Types, Vectors, Calling Function, Missing Data, Data.Frames, Lists, Matrices, Arrays.
	Reading Data Into R: Reading Csvs, Excel Data,Reading From Databases And Extract Data From Websites.
	Control Structures & Loops: If And Else, Switch, If Else, Compound Tests, For Loops, WhileLoops, Controlling Loops.
	Case Study: Loading Data From Mysql Into Rstudio.
	UNIT II:
	Group manipulation: Apply Family, aggregate, plyr, data.table.
	Data Reshaping:cbind, rbind, joins reshape2.
	Strings: paste, sprint, extracting text, regular expressions.
	Math Functions: Cumulative sums and products, minima and maxima, sorting, set operations.
	UNIT III:
	Probability Distributions: Normal Distribution, Binomial Distribution, Poisson Distribution.
	Basics Statistics: Summary statistics, correlation and covariance, t-tests,

	ANOVA. **Linear Models:** Simple Linear Regression, Multiple Regression, Logistics Regression, Poisson Regression.
	UNIT IV: **Nonlinear Models:** Nonlinear Least Squares, Splines, Decision Trees, Random Forests. **Time Series:** Autoregressive Moving Average, Var. **Clustering**: K Means, Pam, Hierarchical Clustering **Plots:** Base Graphics And Ggplot2.
Text books and Reference books	**Text Book(s):** [1]. Jared P. Lander, "R for Everyone, Addison Wesley Data & Analytics Series, Pearson", 2014. [2]. Norman Matloff, "The Art of R Programming, No Strach Press", San Francisco, 2011. **Reference Books:** [1]. Jeffrey Stanton, "An Introduction To Data Science", 2012 [2]. G. Jay Kerns, Introduction to Probability and Statistics using R, First Edition, 2010
E-resources and other digital material	[1]. **Rafael Irizarry, Michael Love, Statistics with R, Harvard University (18, May, 2021). Available: https://www.edx.org/course/statistics-r-harvardx-ph525-1x-1** [2]. Mine Çetinkaya-Rundel, David Banks, Colin Rundel, Merlise a Clyde, Duke University, (18, May, 2021). Statistics with R Specialization. Available: https://www.coursera.org/specializations/statistics

23ITDS1014B – Advanced Java Programming

Course Category:	Programme Elective- I		Credits:		3
Course Type:	Theory		Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	Core Java Programming		Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Understand features of Spring Boot, Spring Framework, Spring cloud and process involved to connect to Java Database Connectivity			
	CO2	Apply concepts of Servlets to develop server side applications			
	CO3	Design web applications with Spring Boot Annotations and connecting to JPA with Spring MVC and Spring Boot			
	CO4	Develop Representational State Transfer services in Spring Boot applications Understand Object Oriented Programming and threads concepts in Java.			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	2		2	2	1
	CO2	2		1	2	
	CO3	2		2	2	2
	CO4	2		2	2	

Course Content	**Unit I** **JDBC**: The concept of JDBC,JDBC Driver Types, JDBC Packages, A Brief Overview Of The JDBC Process, Database Connection, Associating The JDBC/ODBC bridge with the Database, Statement objects, ResultSet. **Java Servlets** :Java Servlets and common gateway interface programming, benefits of using a java servlets, simple java servlet, anatomy of a java servlet, deployment descriptor, reading data from a client, sending data to a client, working the cookies and tracking sessions
	UNIT II: **Getting started with Spring Boot**: Structure, objectives, introduction, features, advantages of Spring Boot, Breaking the monolithic way of developing software, system requirements, setting up of the environment, the 12-factor app, Spring initializer **Developing Spring Boot Application**: Starting with Spring initializer, Build tools, understanding pom file, build.gradel understanding, building an application using Maven and Gradle, understanding the entry PInt class and SpringBootApplication, Bootstrap ApplicationContext
	UNIT III: **Spring Boot Starter Dependencies and Auto-Configuration**: Objectives, Spring Boot Starters, starter dependencies and their configurations,

	understanding auto-configuration.
	Spring Boot Annotations: Java Annotations, existence of spring annotations, Spring and Spring Boot annotations, Stereotype annotations, Spring Boot Annotations, Spring Task execution annotations, Spring profile annotations
	UNIT IV: **Working with Spring Data JPA**: Accessing relational data using JdbcTemplate and Spring data JPA in memory database, Spring data JPA with MySQL, Query methods in Spring data JPA **Micro services: Building RESTful Microservices**: Creating RESTful APIs, Consuming, RESTful APIs. Spring Cloud: Introduction, Features of Spring cloud, Spring Cloud dataflow, features of spring cloud dataflow **Case Study: Deploy Web application into a server using Servelt/Spring Technology**
Text books and Reference books	**Text Book(s):** [1]. James Keogh, "J2Ee: The Complete Reference", 1st Edition, McGraw Hill Education, 2002 [2]. ShagunBakliwal, Hands-on Application Development using Spring Boot, BPB Publications, First Edition, 2022 **Reference Book(s):** [1]. Craig Walls, Spring in Action, Sixth Edition, MEAP Edition, Manning Early Access Program, Version 4, 2021 [2]. Mark Heckler, Spring Boot: Up and Running, O'Reilly Media,2021
E-resources and other digital material	[1]. RangaKaranam, Java Servlets and JSP - Build Java EE(JEE) app in 25 Steps, 04-06-2022 Available: https://www.udemy.com/course/learn-java-servlets-and-jsp-web-application-in25-steps/ [2]. Spring-Official documentation, 04-06-2022 Available:https://spring.io/projects/springboot [3]. Advanced Java Programming by Infinite Skills, 04-06-2022 Available: https://www.udemy.com/advanced-java-programming/ [4]. Derek Parsons , Spring MVC, Spring Boot and Rest Controllers, Available: 04-06-2022, LearnQuest, https://www.coursera.org/learn/spring-mvc-rest-controller [5]. RangaKaranam, Spring Framework Master Class - Java Spring the Modern Way, Available: 04-06-2022 https://www.udemy.com/course/spring-tutorial-for-beginners/

23ITDS1014C –Data Analysis with Python

Course Category:	Programme Elective- I		Credits:		3
Course Type:	Theory		**Lecture-Tutorial-Practice:**		3-0-0
Prerequisites:	Any Programming Language		**Continuous Evaluation:**		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Illustrate the fundamental concepts of Python for data analysis.			
	CO2	Comprehend data by assessing its characteristics, engage in data preprocessing, and data visualization techniques			
	CO3	Analyze various Python packages, including those for mathematical, scientific, and web data analysis.			
	CO4	Evaluate the model development process for data analysis, and performance assessment.			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	1				
	CO2	2	1			
	CO3	2		2	3	
	CO4	2	1		2	

Course Content	**UNIT I:** **Python Fundamentals for Data Analysis** Python data structures, Control statements, Functions, Object Oriented programming concepts using classes, objects and methods, Exception handling, Implementation of user-defined Modules and Package, File handling in python.
	UNIT II: **Introduction to Data Understanding and Preprocessing:** Knowledge domains of Data Analysis, Understanding structured and unstructured data, Data Analysis process, Importing Dataset: Importing and Exporting Data, Basic Insights from Datasets, Cleaning and Preparing the Data: Identify and Handle Missing Values.
	UNIT III: **Data Processing and Visualization** Data Formatting, Exploratory Data Analysis, Filtering and hierarchical indexing using Pandas. Data Visualization: Basic Visualization Tools, Specialized Visualization Tools, Seaborn Creating and Plotting Maps. **Mathematical and Scientific Methods for Data Analysis** Numpy and Scipy Package, Understanding and creating N-dimensional arrays,

	Basic indexing and slicing, Boolean indexing, Fancy indexing, Universal functions, Data processing using arrays, File input and output with arrays.
	UNIT IV: Analyzing Web Data Data wrangling, Web scrapping, Combing and merging data sets, Reshaping and pivoting, Data transformation, String Manipulation, case study for Market Research and Competitive Analysis: Web Scraping for Business Insights, News and Sentiment Analysis: Web Scraping for Financial Markets.
Text books and Reference books	**Text Book(s):** [1]. Chen, D.Y., 2017. Pandas for everyone: Python data analysis. Addison Wesley Professional. [2]. McKinney, W., 2012. Python for data analysis" O'Reilly Media, Inc." [3]. Thareja, R., 2018. Python Programming: Using Problem Solving Approach. Oxford University Press. **Reference Books:** [1]. Brown, T.R., 2023. An Introduction to R and Python for Data Analysis: A Side-By-Side Approach. CRC Press. [2]. Allen Downey ,Jeffrey Elkner ,Chris Meyers,: Learning with Python, Dreamtech Press [3]. David Taieb ,"Data Analysis with Python: A Modern Approach " 1st Edition, Packt Publishing
E-resources and other digital material	[1]. Python Data Science Handbook by Jake VanderPlashttps://jakevdp.github.io/pythondatasciencehandbook/ [2]. DataCamp offers a variety of online courses on Python for data analysis. Some are free, and some require a subscription https://www.datacamp.com/ [3]. Kaggle is a popular platform for data science competitions and offers free datasets, notebooks, and tutorials.https://www.kaggle.com/kernels [4]. Corey Schafer has a comprehensive playlist of Python tutorials that include data analysis topics. https://www.youtube.com/user/schafer5/playlists [5]. Sentdex has a series of videos on Python programming for data analysis and machine learning. https://www.youtube.com/user/sentdex

23ITDS1015A -DATA SCIENCE FOR DECISION MAKING

Course Category:	Programme Elective -II		Credits:		3
Course Type:	Theory		Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	Basic statistics		Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Outline the concept of data driven decision making.			
	CO2	Apply the knowledge of data analysis to solve decision problems.			
	CO3	Identify appropriate courses of action for a given managerial situation whether a problem or an opportunity			
	CO4	Design viable solutions to decision making problems.			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	1	1	2	2	
	CO2	1	1		2	
	CO3	1	1	2		2
	CO4	2	1		2	1

Course Content	**UNIT I:**Fundamentals of Analytics: Introduction to data-driven decision making; generalintroduction to data driven strategy and its importance; use of examples and mini-casestudies to illustrate the role of statistical analysis in decision making.
	UNIT II: Basic Data Analysis: Various types of data that are commonly collected by firms;mean comparisons, andcross tabulation; statistical inferences using chi-square; t-test and ANOVA.
	UNIT III: Experimental Design and Natural Experiments: Issues of design of experiments and internaland external validity; case studies in marketing; economics; and medicine etc.; A-B testing; and circumstances that provide us with "natural" experiments.
	UNIT IV: Decision making tools: Regression analysis and its applications; use of regression output inforecasting; promotional planning and optimal pricing; multivariate cluster analysis; factor analysis decision trees; elastic nets and random forests.
Text books and Reference books	**Text Book(s):** [1]. F.S. Hillier and G.J. Liberman "Introduction to Operations Research" Tata McGrawHill Education Private Limited. [2]. Gregory S. Parnel, Terry A. Bresnick, Steven N. Tani, Eric R. Johnson "Handbook ofDecision Analysis", Wiley.

	Reference Books: **[1].** Emily Moberg and Igor Linkov "Multi-Criteria Decision Analysis: EnvironmentalApplications and Case Studies", CRC Press, Taylor and Francis group. **[2].** Adiel Teixeira de Almeida, EmelAktas, Sarah Ben Amor, João Luis de **[3].** Miranda"Advanced Studies in Multi-Criteria Decision Making", CRC Press.
E-resources and other digital material	**[1].** Data Science For Beginners | Edureka, https://www.youtube.com/watch?v=-ETQ97mXXF0XXX **[2].** Statistics - A Full University Course on Data Science Basics, freeCodeCamp.org https://www.youtube.com/watch?v=xxpc-hpkn28 **[3].** Data Science for Beginners, Google Career Certificates https://www.youtube.com/watch?v=4dlstsqpy84

23ITDS1015B -CLOUD DATA ENGINEERING

Course Category:	Programme Elective -II		Credits:	3
Course Type:	Theory		Lecture-Tutorial-Practice:	3-0-0
Prerequisites:	-		Continuous Evaluation:	40
			Semester end Evaluation:	60
			Total Marks:	100
Course Outcomes	Upon successful completion of the course, the student will be able to:			
	CO1	Summarize the needs to migrate databases onto the cloud systems		
	CO2	Identify data management in the cloud system to minimize risks of data loss and improper data handling.		
	CO3	Apply the cloud features to protect systems on the network		
	CO4	Analyze the impact of using cloud data systems and its migration		

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2- Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1			2		
	CO2		1	2		
	CO3	1	1	1		
	CO4		1	2		

Course Content	UNIT I: **Data Management :**Introduction to Data Management, Safeguarding data , Making the most of data, Models of data management - The silo approach, The federated approach. **Modern Data Management with the Cloud** :Understanding Cloud Services, Cloud Data Management Basics, Centralized access and visibility, Centralized backups and archives, Centralized searches, Centralized governance, Centralized analysis and risk awareness, Cloud-Native Technology.
	UNIT II: **Backup And Disaster Recovery :**Traditional Backups and Their Limitations, The Cloud-Native Backup, Types of Data Protection :Protecting Servers, Disaster readiness, Protecting user data, Six Backup Musts. **Accessing And Analyzing Data :**Governance, The Traditional Approach to Governance, The Cloud Approach to Governance.
	UNIT III: **Minimizing Data Risks :**Guarding Against Malware, Checking for Data Leaks, Monitoring for Anomalies, Ensuring Legal and Regulatory Compliance, Data Privacy and Access, Data Collection and Retention**.** **Ten Reasons To Use Cloud Data Management :** Assurance of Comprehensive Data Collection, Simplifies Backup and Recovery, Works Across Locations Worldwide, Easy to Analyze Data for Trends, Makes Malware/Ransomware Recovery Easier, Early Warning for Potential Data Access Anomalies, Ensures Compliance with Regulations, Makes E-discovery Quicker and Easier,

	Invisible to End-users, Saves Money Compared to Other Options
	UNIT IV: **Databases in the Cloud :**High level effects of moving to the cloud, Self-Managed Versus Managed Databases,Cloud Native Databases,Types of Managed Databases,Role of the DBA in a Managed Database. **Moving Databases to the Cloud :**Planning, Factors in a Migration, Major Migration Tasks, Readiness Assessment, Checking for Incompatibilities, Data Movement, Migrating the Database, Migrating Applications, Post-Migration Checks.
Text books and Reference books	**Text Book(s):** [1]. Wendy A. Neu, VladVlasceanu, Andy Oram& Sam Alapati, "An Introduction to Cloud Databases", O'Reilly Media, Inc., 2019 [2]. FaitheWempen, "Cloud Data Management for Dummies", John Wiley & Sons, Inc, 2017. **Reference Books:** [1]. Liang Zhao , SherifSakr , Anna Liu , AthmanBouguettaya, "Cloud Data Management", Springer Cham,2014.
E-resources and other digital material	[1]. Courseera, "Database instance on Cloud", 2020 https://www.coursera.org/lecture/sql-data-science/how-to-create-a-database-instance-on-cloud- [2]. Craig Stedman, Industry Editor, "Cloud DBMS guide", 2022 https://www.techtarget.com/searchcloudcomputing/definition/cloud-database

23ITDS1015C-CYBER SECURITY&FORENSICS

Course Category:	Program Elective-II		Credits:		3
Course Type:	Theory		Lecture-Tutorial-Practice:		3-0-0
Prerequisites:			Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Categorize various types of attacks in Information security			
	CO2	Apply data leakage prevention, protection and security policies on data			
	CO3	Explore the role of Digital Forensics and its readiness planning in investigation Process			
	CO4	Analyze First Responder Procedure through Computer Forensics Investigation Process			
Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO4	PO5
	CO1		1			1
	CO2	1			2	
	CO3	3		1	1	1
	CO4		2	2		
Course Content	**UNIT I:** **Information Security and Threats:** Introduction – Information Security, Information Assets & Threats - Threats to Information Assets, Types of Attacks, Types of Virus, Types of Worms, Types of Trojans, Network Attacks, Common Vulnerabilities and Exposures (CVE). **Fundamentals of Information Security:** Elements of information security – Network Security, Application Security, Communications Security. Principles and concepts – data security – Critical Information Characteristics, Information States,Prevention Vs Detection, Types of controls – Access Control Models					
	UNIT II: **Data Leakage and Prevention:** Introduction to Data Leakage, Organisational Data Classification, Location and Pathways, Content Awareness, Content Analysis Techniques, Data Protection **Network Sniffers and Injectors**–Sniffers Overview, Tcpdump, Wireshark, Ettercap					
	Unit III: **Introduction To Digital Forensic**: Introduction, Evolution of Computer Forensics, Stages of Computer Forensics Process, Benefits of Computer Forensics, Uses of Computer Forensics, Objectives of Computer Forensics, Role of Forensics Investigator **Forensics Readiness:** What Is Forensics Readiness, Goals of Forensic					

	Readiness, Benefits of Forensic Readiness, Steps For Forensic Readiness Planning
	Unit IV: **Computer Forensics Investigation Process:** Introduction To Computer Crime Investigation, Assess The Situation, Acquire The Data, Analyze The Data, Report The Investigation **Digital Evidence And First Responder Procedure:** Digital Evidence, First Responder Toolkit, Issues Facing Computer Forensics, Types of Investigation, Techniques of Digital Forensics
Text books and Reference books	**Text Book(s):** [1].Student Handbook – Security Analyst, NASSCOM,2015 [2]. Anti-Hacker Tool Kit (Indian Edition) Fourth Edition by Mike Shema, Publication Mc Graw Hill,2014 [3].Dr.JeetendraPande Dr. Ajay PrasadUttarakhand Open University, "Digital Forensics" Haldwani publishers, 2016 **Reference Books:** [1]. Nina Godbole and SunitBelpure, "Cyber Security Understanding Cyber Crimes", Computer Forensics andLegalPerspectives Publication, Wiley,2012 [2]. Cyber Security, Understanding cyber crimes, computer forensics and legal perspectives, Nina Godbole,SunitBelapure, Wiley Publications, Reprint 2016
E-resources and other digital material	[1] Prof.V.Kamakoti ,Professor, Introduction to Information security, IIT Madras ,Jan-Mar 2015 ,https://nptel.ac.in/courses/106106129 [2] Prof.ChesterRebeiro , Professor ,Secure System Engineering ,IIT Madras, Jan-Mar 2023 ,https://nptel.ac.in/courses/106106199 [3] Sanjay Goel, Associate Professor , Introduction to Cybercrime and Fundamental Issues, Sep 2022 ,https://in.coursera.org/lecture/cyber-conflicts/introduction-to-cybercrime-and-fundamental-issues-xndSq [4]https://www.bt.com.au/professional/knowledge-centre/business-resources/business-development/targeted-malware-attacks.html [5] RavindraSavaram, "CyberArc tutorial" January 2023 https://mindmajix.com/cyberark-tutorial

23ITDS1051-ADVANCED DATA STRUCTURES AND ALGORITHMS LAB

Course Category:	Laboratory - I		Credits:	1.5
Course Type:	Laboratory		**Lecture-Tutorial-Practice:**	0-0-3
Prerequisites:	Any programming language		**Continuous Evaluation:**	40
			Semester end Evaluation:	60
			Total Marks:	100
Course Outcomes	Upon successful completion of the course, the student will be able to:			
	CO1	Implement operations on tree data structures.		
	CO2	Perform operations on balanced data structures		
	CO3	Apply graph data structure to solve real world problems		
	CO4	Design an optimal solution using appropriate data structures and design techniques		

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	2				
	CO2	2				
	CO3	2				
	CO4	3		2		

Course Content	**Week 1: Analyze the time complexity of Algorithms**
	a. Implement the logic to find the square root of a given and analyze its time complexity and write another logic with better time complexity.
	b. Any design experiment on time and space complexity analysis.
	Week 2: Binary search tree and applications
	a. Implementation of Binary search tree operations.
	b. Application on Binary Search trees
	Week 3: AVL and applications
	a. Insert and delete operations on AVL-tree
	b. Application on AVL trees
	Week 4:B- tree and applications
	a. Insert and delete operations on B-tree
	b. Application on B-trees
	Week 5: Basic Graph Algorithms
	a. Create a graph with insertion and deletion of nodes and edges
	b. Graph traversal techniques-DFS and BFS
	Week 6: Shortest Path Algorithms
	a. Dijkstra's Algorithm

	b. All-pairs shortest path
	c. Application on Shortest Path Algorithms
	
	Week 7: Minimum Cost Spanning Trees
	    a. Prim's Algorithm
	    b. Kruskal's Algorithm
	
	Week 8: Greedy Algorithms
	    a. Optimal Merge Patterns/Optimal Storage on tapes
	    b. Application problem with Greedy Algorithm
	
	Week 9: Dynamic Programming
	    a. Multi-stage graphs/Optimal Binary Search Trees
	    b. Application problem with dynamic programming
	
	Week 10: Design experiments/scenario based problem solving usingAdvancedData structures
	
	Week 11: Design experiments/scenario based problem solving usingAlgorithm Design Techniques

Text books and Reference books	**Text Book(s):** **[1].** Mark Allen Weiss, Data Structures and Algorithm Analysis in C++, 4th Edition, Pearson Publication **[2].** Ellis Horowitz, SartajSahni and S. Rajasekharan, Fundamentals of Computer Algorithms, 2nd Edition, , Universities Press. **Reference Books:** **[1].** YedidyahLangsam, Moshe J. Augenstein and Aaron M. Tenenbaum, "Data Structures using C and C++", 2nd edition, Pearson Education, 1999. **[2].** Data Structures: A Pseudocode Approach with C, 2nd Edition, R. F. Gilberg and B. A. Forouzan, Cengage Learning
E-resources and other digital material	**[1].** Erik Demaine, Advanced Data Structures, [MIT- OpenCourseWare]. (26, May, 2021). Available: http://ocw.mit.edu/ **[2].** Dr. Naveen Garg, Department of Computer Science & Engineering ,IIT Delhi, Lecture Series on Data Structures and Algorithms [NPTEL], (26,May,2021) Available: https://nptel.ac.in/courses/106/102/106102064/ **[3].** Data Structures and applications on, [Geeksforgeeks], (25, May, 2021) Available: https://www.geeksforgeeks.org/data-structures/ **[4].** Data Structures and challenges [Hacker rank], (25,May,2021) Available: https://www.hackerrank.com/domains/data-structures

23ITDS1052A – STATISTICS WITH R LAB

Course Category:	Laboratory - II		Credits:	1.5
Course Type:	Laboratory		Lecture-Tutorial-Practice:	0-0-3
Prerequisites:	Any programming language and Basic Mathematics		Continuous Evaluation:	40
			Semester end Evaluation:	60
			Total Marks:	100
Course Outcomes	Upon successful completion of the course, the student will be able to:			
	CO1	Interpret different types of data manipulation and group manipulation operations		
	CO2	Apply data visualizations tools to displaypatterns and insights of data.		
	CO3	Build classification and regression models in R		
	CO4	Develop solutions to data analysis problems using statistical techniques		

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3-High)		PO1	PO2	PO3	PO 4	PO 5
	CO1			2		
	CO2					3
	CO3	1			3	
	CO4	1		2		

Course Content	**week-1:** Performing basic R commands Write a R-program to check whether a number is even or odd Write the R-program to check whether number is palindrome ornot. Write a R-program to print Fibonacci series
	week-2: Implement function calls and commands Implement math functions in R studio Implement string manipulation functions melt and cast
	week-3: Extracting csv file from web browserusing URL. Extracting data from mysql Perform the operations on list, arrays and matrix.
	week-4: Implement group manipulation functions using Apply family Implement group manipulation functions using aggregate ,plyr functions
	week-5: Implement Data Reshaping concepts like cbind(), rbind() in R studio. Implement different types of joins in Rstudio. Implement basic summary statistics concept in R studio
	week-6: Implement basic summary statistics concept in Rstudio

	Implement different types of math functions in R.
	week-7: Analyze the correlation and covariance for the different attributes for given data set using R
	week-8: Implement statistical distribution concepts Normal, Binomial, Poisson Distributions on the given application using R.
	week-9: Implement simple linear, multiple linear, poisons and logistic regressions on the given application using R.
	week-10: Implement k-means, k-mediods and hierarchical techniques on the given application using R.
	week-11: Implement the concept of statistical graphs in Rstudio.
	Week-12: Case study: Implement the test case that predicts the possibility of success of launching of new news paper when the sales of the existing sales papers are given**.**
Text books and Reference books	**Text Book(s):** [1].Jared P. Lander, "R for Everyone, Addison Wesley Data &Analytics Series, Pearson", 2014Norman Matloff, "The Art of R Programming, No Strach Press", San Francisco, 2011 **Reference Books:** [1].Jeffrey Stanton, "An Introduction To Data Science", 2012 [2].G. Jay Kerns, Introduction to Probability and Statistics using R, First Edition, 2010
E-resources and other digital material	[1].Dr.Shalabh is a Professor of Statistics at IIT Kanpur, NPTEL course Descriptive Statistics with R Software. Available: noc19-ma14-Introduction - Descriptive Statistics with R Software - YouTube [2].Rafael Irizarry, Michael Love, Statistics with R, Harvard University (18, May, 2021). Available: https://www.edx.org/course/statistics-r-harvardx-ph525-1x-1 [3].Mine Çetinkaya-Rundel, David Banks, Colin Rundel, Merlise A Clyde, Duke University, (18, May, 2021). Statistics with R Specialization. Available: https://www.coursera.org/specializations/statistics

23ITDS1052B - ADVANCED JAVA PROGRAMMING LAB

Course Category:	Laboratory-II		Credits:	1.5
Course Type:	Lab		Lecture-Tutorial-Practice:	0-0-3
Prerequisites:	Core Java		Continuous Evaluation:	40
			Semester end Evaluation:	60
			Total Marks:	100

Course Outcomes	Upon successful completion of the course, the student will be able to:	
	CO1	Implement Java Database Connectivity Application Programming Interface to connect to relational databases
	CO2	Build server side applications to interact with server using Java Servlets
	CO3	Implement dependency injection and inversion of control to solve problems in Spring Boot.
	CO4	Create Spring Boot applications to solve real world problems that uses Representational State Transfer services

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3-High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	2		2		
	CO2	2		2		3
	CO3	2		2	3	
	CO4	2		2		

Course Content	**Week-1:** **Create JDBC programs to connect to relational databases** for the following operations
	  a. To Insert data
	  b. To query the data
	  c. To analyse the data using SQL aggregate operations

	week-2: Connect to a database using different type of Statement Interfaces and process the results using ResultSet Implementation of CRUD operations on a relational database
	Week-3: Create Server side applications using Java Servlets
	Week 4: Servlet programs on sessional tracking using a. Cookies b. Sessions
	Week-5: Creation of Spring programs via SpringBootApplication and Spring Initializerin Spring

	Week-6: Implementation of 12-factor App in Spring Boot
	Week-7: Implement dependency injection into a program in Spring Boo
	Week-8: Use of annotations in developing applications in Spring Boot
	Week-9: Accessing of relational databases via JDBC and JPA
	Week-10: Implement RESTFul Services in Spring Boot
	Week-11: Build an application using Spring Boot Representational State Transfer for any usecase
	Week-12: **Case Studies** 1**.** Develop web applications using Java Servlets 2. Web applications that handles the sessions via session tracking 3. Develop Spring Boot applications to real world problems 4. Make use of Representational State Transfer in building applications in Spring Boot
Text books and Reference books	**Text Book(s):** **[1].** James Keogh, "J2Ee: The Complete Reference", 1st Edition, McGraw Hill Education, 2002 **[2].** ShagunBakliwal, Hands-on Application Development using Spring Boot, BPB Publications, First Edition, 2022 **Reference Book(s)**: **[1].** Craig Walls, Spring in Action, Sixth Edition, MEAP Edition, Manning Early Access Program, Version 4, 2021 **[2].** Mark Heckler, Spring Boot: Up and Running, O'Reilly Media,2021
E-resources and other digital material	**[1].** RangaKaranam, Java Servlets and JSP - Build Java EE(JEE) app in 25 Steps, 04-06-2022 Available: https://www.udemy.com/course/ learn - java-servlets-and-jsp-web-applicationin-25-steps/ **[2].** Spring-Official documentation, 04-06-2022 Available: https://spring.io/projects/springboot **[3].** Advanced Java Programming by Infinite Skills, 04-06-2022 Available: https://www.udemy.com/advanced-java-programming/ **[4].** Derek Parsons , Spring MVC, Spring Boot and Rest Controllers, Available: 04-06-2022, LearnQuest, https://www.coursera.org /learn/spring-mvc-rest-controller **[5].** RangaKaranam, Spring Framework Master Class - Java Spring the Modern Way, Available: 04-06-2022 https://www.udemy.com/ course/spring-tutorial-for-beginners/

23ITDS1052C-DATA ANALYSIS WITH PYTHON LAB

Course Category:	Laboratory-II		Credits:	1.5
Course Type:	Lab		Lecture-Tutorial-Practice:	0-0-3
Prerequisites:	Any Programming Language		Continuous Evaluation:	40
			Semester end Evaluation:	60
			Total Marks:	100
Course Outcomes	Upon successful completion of the course, the student will be able to:			
	CO1	Demonstrate competence in decision control, string handling, list manipulation, and object-oriented design.		
	CO2	Apply data analysis skills to glean insights from diverse datasets.		
	CO3	Analyze web data critically to derive actionable business insights.		
	CO4	Create innovative data solutions, employing advanced processing and visualization techniques.		
	CO5	Evaluate data quality and reliability critically, cultivating discernment and data-driven decision-making for informed outcomes.		

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	2				
	CO2	2			2	
	CO3	3		2	2	
	CO4	3		2	2	3
	CO5	3		2	2	3

Course Content

Week I:Python programs on Decision Control Statements
1. Implement programs on and conditional branchingstatements
2. Understand Scope of a variable and Use globalstatement
3. Differentiate fruitful and void functions through a casestudy
4. Apply recursive and Lambdafunctions

WeekII:Solve the problems using Strings and understanding the methods
1. Apply string formattingoperator
2. Use built in string methods, functions and regularexpressions
3. Define a list and write programs to access and modify elements of alist
4. Practice basic list operations, methods

Week III:Implement the Python Classes and Objects to address the real world scenarios
1. Define classes and objects using python for the real worldscenario
2. Defining constructors and usingSelf
3. Understand public and privatemembers

4. Practice calling class methods from anotherclass
5. Write built in functions to check, get, set and deleteattributes

Week IV:Importing and Exploring Datasets
1. Import a dataset (e.g., a CSV file) and display its first few rows using Pandas.
2. Calculate summary statistics (mean, median, standard deviation) for numerical columns in the dataset from kaggle.
3. Identify and display data types of each column in the dataset.
4. Check for missing values in the dataset and create a summary report.

Week V:Data Cleaning
1. Handle missing values by imputing them with the mean, median, or mode.
2. Detect and remove duplicate rows from the dataset.
3. Explore and clean columns with inconsistent data formatting (e.g., dates with different formats).
4. Create a cleaned and well-structured dataset for further analysis.

Week VI:Data Formatting and EDA
1. Convert date columns into a consistent format (e.g., 'YYYY-MM-DD') using Pandas datetime functions.
2. Clean text data by converting all strings to lowercase and removing leading/trailing whitespaces.
3. Create a histogram to visualize the distribution of a numerical variable.
4. Generate a scatter plot to explore the relationship between two numerical variables.

Week VII:Advanced Data Visualization
1. Create a Seabornheatmap to visualize the correlation matrix of numerical variables.
2. Build a pair plot to visualize pairwise relationships among multiple numerical variables.
3. Construct a box plot to explore the distribution of a numerical variable across different categories.
4. Use Seaborn'sFacetGrid for advanced custom visualizations based on dataset characteristics.

Week VIII:Mathematical and Scientific Applications
1. Create a 2D Numpy array, perform basic indexing, and extract specific rows/columns.
2. Apply Boolean indexing to filter rows based on a condition.
3. Use universal functions (ufuncs) for element-wise mathematical operations.
4. Utilize Scipy for basic scientific computing tasks like integration or solving linear equations.

Week IX:Data Processing with Numpy
1. Aggregate data by grouping and calculating summary statistics.
2. Filter data using logical conditions and create subsets of the dataset.
3. Apply statistical functions (e.g., mean, median, standard deviation) to Numpy arrays.
4. Save and load Numpy arrays from external files (e.g., CSV, binary files).

	Week X: Analyzing Web Data: Web Scraping Basics
	  1. Scrape headlines from a news website and store them in a Pandas DataFrame.
	  2. Extract links and publication dates from the scraped articles.
	  3. Export the scraped data to a CSV file for further analysis.

	Week XI:Combining and Transforming Web Data
	  1. Scrape data from multiple web sources with different structures (e.g., tables and JSON).
	  2. Combine data from different sources into a single DataFrame.
	  3. Reshape and pivot the data to create meaningful insights.
	  4. Perform data transformation tasks like string manipulation and data type conversion.

	Week XII:Market Research and Sentiment Analysis
	  1. Scrape financial news articles from multiple sources.
	  2. Preprocess text data by removing stopwords and performing tokenization.
	  3. Perform sentiment analysis on the text using NLP libraries (e.g., NLTK or spaCy).
	  4. Visualize sentiment trends over time using line charts or bar plots.

Text books and Reference books	**Text Book(s):** [3]. Chen, D.Y., 2017. Pandas for everyone: Python data analysis. Addison-Wesley Professional. [4]. McKinney, W., 2012. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc." **Reference Books:** [1]. Allen Downey ,Jeffrey Elkner ,Chris Meyers,: Learning with Python, Dreamtech Press [2]. David Taieb ,"Data Analysis with Python: A Modern Approach " 1st Edition, Packt Publishing [3]. Brown, T.R., 2023. An Introduction to R and Python for Data Analysis: A Side-By-Side Approach. CRC Press.
E-resources and other digital material	[1]. Python Data Science Handbook by Jake VanderPlashttps://jakevdp.github.io/pythondatasciencehandbook/ [2]. DataCamp offers a variety of online courses on Python for data analysis. Some are free, and some require a subscription https://www.datacamp.com/ [3]. Kaggle is a popular platform for data science competitions and offers free datasets, notebooks, and tutorials. Https://www.kaggle.com/kernels [4]. Corey Schafer has a comprehensive playlist of Python tutorials that include data analysis topics. Https://www.youtube.com/user/schafer5/playlists [5]. Sentdex has a series of videos on Python programming for data analysis and machine learning.https://www.youtube.com/user/sentdex

SEMESTER II

23ITDS2001-BIGDATA FRAMEWORK FOR DATA SCIENCE

Course Category:	Programme Core–IV	Credits:		3
Course Type:	Theory	Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	Database Management Systems	Continuous Evaluation:		40
		Semester end Evaluation:		60
		Total Marks:		100

Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Summarize Big Data Characteristics, Hadoop, Hive, Hdfs And Map Reduce Architectures.			
	CO2	Experiment With Nosql Databases To Process Unstructured And Semi Structured Data.			
	CO3	Apply Pig Latin, Hive Scripts And Map Reduce Programming On Real Time Applications.			
	CO4	Perform In-Memory Data Analytics With Spark And Spark Streaming.			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	2		1		3
	CO2	2		1		3
	CO3	2		1		3
	CO4	2		1		3

Course Content	**UNIT I:** **INTRODUCTION TO BIG DATA:** Big Data-Definition, Characteristics of Big Data (Volume, Variety, Velocity), Data in the Warehouse and Data in Hadoop, Importance of Big Data. **INTRODUCTION TO HADOOP:** Data, Data Storage and Analysis, Comparison with other systems: Rdbms, A Brief History of Hadoop, The Hadoop Ecosystem, Hadoop Releases.
	UNIT II: **NoSQL:** Introduction to NOSQL, Types of NoSQL Databases, Advantages of NoSQL databases, SQL versus NoSql. **NoSQL databases:** Introduction to MongoDB, Data types in MongoDB, MongoDB query language. **Hadoop Distributed File System:** The Design of HDFS, HDFS Concepts, Blocks, Namenodes and Datanodes, Anatomy of a File Read and Anatomy of a File Write.
	UNIT III: **Map Reduce**–A Weather Dataset, Data Format, Analyzing the Data with Unix Tools, Analyzing the Data with Hadoop, Map and Reduce, Java Map Reduce. **Pig-**Execution Types, Running Pig Programs, Pig Latin Editors, Comparison with databases, Pig Latin, Functions, Data Processing Operators.

	UNIT IV: **Hive**-Hive Services, Comparison With Traditional Databases, Hiveql, Tables, Querying Data. **Spark-**Introduction To Data Analytics With Spark, Spark Stack, Programming with RDDS, Working With Key/Value Pairs And Spark Sql, Spark Streaming.
Text books and Reference books	**Text Book(s):** [1].Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, "Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data" 1st Edition, TMH,2012. [2].Tom White, Hadoop, "The Definitive Guide", 3rd Edition, O'Reilly Publications, 2012. [3].Seema Acharya, SubhashiniChellappan, Big Data and Analytics, Wiley Publishers. **Reference Books:** [1].Holden Karau, Andy Konwinski, Patrick Wendell, MateiZaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly Media, Inc.
E-resources and other digital material	[1].Big Data Use cases for Beginners | Real Life Case Studies | Success Stories https://www.youtube.com/watch?v=HHR0-iJp2sM [2].Alexey Grishchenko, Hadoop Vs Mpp, Https://0x0fff.Com/Hadoop-Vs-Mpp/ [3].Random Notes On Bigdata- Slideshare: Available www.Slideshare.Net/Yiranpang/Random-Notes-On-Big-Data-26439474 [4].Introduction To Big Data, IlkayAltintas,Amarnath Gupta, Https://Www.Coursera.Org/Learn/Big-Data-Introduction?Specialization=Big-Data

23ITDS2002-DEEP LEARNING

Course Category:	Program Core-V		Credits:		3
Course Type:	Integrated Course		Lecture-Tutorial-Practice:		2-0-2
Prerequisites:	Machine Learning		Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	summarize basic concepts of neural networks, back propagation, Attention mechanisms			
	CO2	Apply ANN, CNN, Auto encoders and GANs on image processing applications			
	CO3	Design a suitable RNN model for time series applications			
	CO4	Create a suitable intelligent model for the given application			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1			1		
	CO2	2		1	2	
	CO3	2		1	2	
	CO4	3		2	3	

Course Content	UNIT I: **The Neural Network**: Building Intelligent Machines , The Limits of Traditional Computer Programs , The Mechanics of Machine Learning , The Neuron , Expressing Linear Perceptrons as Neurons , Feed-Forward Neural Networks, Linear Neurons and Their Limitations , Sigmoid, Tanh, and ReLU, Softmax output layers **Training Feed-Forward Neural Networks**: Gradient Descent , The Delta Rule and Learning Rates , Gradient Descent with Sigmoidal Neurons, The Backpropagation Algorithm , Stochastic and Minibatch Gradient Descent , Test Sets, Validation Sets, and Overfitting, Preventing Overfitting in Deep Neural Networks
	UNIT II: **Convolutional Neural Networks**: Neurons in Human Vision ,The Shortcomings of Feature Selection, Vanilla Deep Neural Networks, Filters and Feature Maps, Full Description of the Convolutional Layer, Max Pooling, Full Architectural Description of Convolution Networks **Embedding and Representation Learning**: Learning Lower-Dimensional Representations,Principal Component Analysis, Motivating the Autoencoder Architecture, Denoising, Sparsity in Autoencoders

	UNIT III: **Sequence Modeling: Recurrent and Recursive nets:** Unfolding Computational Graphs, Recurrent neural networks, Bidirectional RNNS,Encoder-Decoder sequence-to –sequence architectures, Deep Recurrent networks, Recursive neural networks. **The Challenge of Long-Term Dependencies:** Echo State Networks, Leaky Units &Other strategies for multiple timescales, The Long Short-Term memory
	UNIT IV: **Advanced Topics in Deep Learning:**Introduction, Attention Mechanisms, Recurrent Models of Visual Attention, Attention Mechanisms for Machine Translation **Generative Adversarial Networks:**Training a Generative Adversarial Network, Using GANs for Generating Image Data, Conditional Generative Adversarial Networks, Limitations of Neural Networks
	Content Beyond: Introduction to Transformer Neural Network, Generative AI
Text books and Reference books	**Text Book(s):** **[1].** Nikhil Buduma, Nicholas Locascio, "Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms", O'Reilly Media, 2017 **[2].** Ian Goodfellow, YoshuaBengio, Aaron Courville, "Deep Learning(Adaptive Computation and Machine Learning series",MIT Press, 2017 **[3].** Charu C. Aggarwal, Neural Networks and Deep Learning, c Springer International Publishing AG, part of Springer Nature 2018, ISBN 978-3-319-94462-3 ISBN 978-3-319-94463-0 (eBook) **Reference (Book)s:** **[1].** Li Deng and Dong Yu, "Deep learning Methods and Applications", Now publishers, 2013 **[2].** Michael Nielsen,"Neural Networks and Deep Learning", Determination Press 2015 **[3].** Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017; 30.
E-resources and other digital material	**[1].** MiteshKhapra, "Deep Learning", Sep 20, 2018, https://www.youtube.com/ watch?v=4TC5s_xNKSs&list=PLH-xYrxjfO2VsvyQXfBvsQsufAzvlqdg9 **[2].** AfshineAmidi and ShervineAmidi ,"Deep Learning cheat sheets for Stanford's CS 230", 2018, https://github.com/afshinea/stanford-cs-230-deep-learning **[3].** YoshuaBengio, Deep learning: "Theoretical Motivations, Canadian Institute for Advanced Research", 2015 http://videolectures.net/deeplearning2015_bengio_theoretical_motivations/ **[4].** Https://Synthesis.Ai/2023/12/04/Generative-Ai-Part-0-Background-On-Transformers/ **[5].** Geoffrey Hinton's GoogleTech Talk,"Recent developments on Deep Learning" March 2010, https://www.youtube.com/watch?v=VdIURAu1-aU

23ITDS2003-DATA VISUALIZATION AND INTERPRETATION

Course Category:	Programme Core – VI		Credits:		3
Course Type:	Theory		Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	Data Structures		Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Articulate objectives of Data Visualization and techniques			
	CO2	Analyze data to create a visualization for various real-time applications			
	CO3	Develop programs and map visual layouts & graphical properties.			
	CO4	Create and publish visualizations that enable clear interpretations of big, complex and real world data			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	`			1	1
	CO2	1				2
	CO3			2		2
	CO4	2			3	3

Course Content	UNIT I: The Context of Data Visualization: Visualization as a discovery tool, The bedrock of visualization knowledge, Defining data visualization, Visualization skills for the masses, the data visualization methodology. Setting the Purpose and Identifying Key Factors: Establishing intent – the visualization's function, Establishing intent – the visualization's tone, Key factors surrounding a visualization project, The " eight hats" of data visualization design
	UNIT II: Conceiving and Reasoning Visualization Design Options: Data visualization design is all about choices, The visualization anatomy – data representation, The visualization anatomy – data presentation Taxonomy of Data Visualization Methods: Data visualization methods, Choosing the appropriate chart type, Assessing hierarchies and part-to-whole relationships
	UNIT III: Constructing and Evaluating Your Design Solution: For constructing visualizations, technology matters, The construction process, Approaching the finishing linePost-launch evaluation. Case Studies on real-time applications.
	UNIT IV: An Introduction to Connecting to Data: An Introduction to Connecting to Data in Tableau, Shaping Data for Use with Tableau, Getting a Lay of the Land: Tableau Terminology, View the Underlying Data, View the Number of Records, Dimension Versus Measure, What Is a Measure? What Is a Dimension? Discrete Versus Continuous Five Ways to Make a Bar Chart/An

	Introduction to Aggregation: Five Ways to Create a Bar Chart in Tableau An Introduction to Aggregation in Tableau, Line Graphs, Independent Axes, and Date Hierarchies, How to Make a Line Graph in Tableau, Independent Axes in Tableau, Date Hierarchies in Tableau, Marks Cards, Encoding, and Level of Detail, An Explanation of Level of Detail, An Introduction to Encoding, Label and Tooltip Marks Cards.
Text books and Reference books	**Text Book(s):** [4]. Andy Kirk, "Data Visualization: a successful design process", Packt Publishing (26 December 2012) [5]. Ryan Sleeper, Practical Tableau, O'Reilly Media, Inc. April 2018. **Reference Books:** [1]. Chakrabarti, S,"Mining the web: Discovering knowledge from hypertext data ",Morgan Kaufman Publishers, 2003. [2]. Ben Fry, Vilisualizing data, Sebastopo,O'Reily, 2007.
E-resources and other digital material	[1]. Dr. Gauravdixit,department of management studies, indian institute of technology,Roorkee: https://nptel.ac.in/courses/110107092/7,2017 [2]. Padammarcus, and eugenewu. Res.6-009 how to process, analyze and Visualize data. January iap 2012. Massachusetts institute of technology: mit open Courseware, https://ocw.mit.edu.,2012 [3]. Prof.shankarnarasimhan,ragunathan, rengasamy,iit madras data Visualization in r basicgraphics,Https://nptel.ac.in/courses/106106179/11,2016 **[4].** Statistics and visualization for data analysis and inference, dr. Ed vul, Dr. Mike frank, massachusetts institute of technology,https://ocw.mit.edu/resources/res-9-0002-statistics-and-visualization-fordata- analysis-and-inference-january-iap-2009/, 2009.

23ITDS2014A- BUSINESS ANALYTICS AND MODELLING

Course Category:	Programme Elective– III	Credits:		3
Course Type:	Theory	Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	Applied Machine Learning	Continuous Evaluation:		40
		Semester end Evaluation:		60
		Total Marks:		100

Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Understand the foundational concepts in business analytics, encompassing the evolution, scope, and models.			
	CO2	Analyze and model probability distributions, enabling them to make informed decisions based on various types of data.			
	CO3	Develop appropriate forecasting techniques, demonstrating an ability to analyze time series data and implement statistical models for accurate predictions.			
	CO4	Integrate optimization methods and decision analysis for solving complex business problems by applying linear optimization and decision-making strategies.			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3-High)		PO 1	PO 2	PO3	PO4	PO5
	CO1	1		2	2	2
	CO2	1			2	2
	CO3	2			2	2
	CO4	3		2	3	3

Course Content	UNIT I: **Foundations of Business Analytics** Introduction, Evolution of Business Analytics, Scope of Business Analytics, Data for Business Analytics, Models in Business Analytics, Problem Solving with Analytics.
	UNIT II: **Descriptive Analytics:** **Probability Distributions and Data Modeling**: Basic concepts of probability, Random variables and probability distribution, Discrete Probability Distributions , Continuous Probability Distributions, Random Sampling from Probability Distributions, Data Modeling and Distribution Fitting. **Sampling and Estimation:** Statistical Sampling, Sampling Error, Sampling Distributions.
	UNIT III: **Predictive Analytics:** **Forecasting Techniques** - Qualitative and Judgmental Forecasting, Statistical

	Forecasting Models, Forecasting Models for Stationary Time Series, Forecasting Models for Time Series with a Linear Trend, Forecasting Time Series with Seasonality, Selecting Appropriate Time-Series-Based Forecasting Models , Regression Forecasting with Causal Variables. **Monte Carlo Simulation and Risk Analysis**- Spreadsheet Models with Random Variables, Monte Carlo Simulation Using Analytic Solver Platform, New-Product Development Model, Newsvendor Model.
	UNIT IV: **Prescriptive Analytics:** **Applications of Linear Optimization-**Process Selection Models, Solver Output and Data Visualization, Blending Models, Portfolio Investment Models, Transportation Models. **Decision Analysis**: Formulating Decision Problems, Decision Strategies without Outcome Probabilities, Decision Strategies with Outcome Probabilities, Decision trees, The value of information, Utility and decision making.
Text books and Reference books	**Text books:** [1] James Evans, "Business Analytics, Second Edition, Pearson Publications, 2017. Reference Books: [1] U. Dinesh Kumar, "Business Analytics - The Science of Data Driven Decision Making", First Edition, Wiley Publications, 2017. [2] SeemaAcharya R N Prasad, "Fundamentals of Business Analytics", 2nd Edition, Wiley Publications, 2016
E-resources and other digital material	[1] Cody Baldwin, "Introduction to Business Analytics", Feb 2016. https://www.youtube.com/channel/UCHPHVCq_Giziio_y8QEcHyA [2] ACADGILD, "Business Analytics for Beginners", 2016 https://www.youtube.com/watch?v=an9PXNtTSSc

23ITDS2014B-IMAGE AND VIDEO ANALYTICS

Course Category:	Program Elective-III	Credits:	3
Course Type:	Theory	Lecture-Tutorial-Practice:	3-0-0
Prerequisites:		Continuous Evaluation:	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes		Upon successful completion of the course, the student will be able to:
	CO1	Illustrate the principles and techniques of digital image in applications related to digital imaging system
	CO2	Understand various image preprocessing techniques and their significance.
	CO3	Analyze various standard deep learning networks for real time applications.
	CO4	Understand the fundamentals of digital video processing

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1					1
	CO2				1	1
	CO3	2				2
	CO4	1			1	1

Course Content	**UNIT I:** **INTRODUCTION** Computer Vision – Image representation and image analysis tasks – Image representations -digitization – properties – color images – Data structures for Image Analysis – Levels of image data representation – Traditional and Hierarchical image data structures.
	UNIT II: **IMAGE PRE-PROCESSING** Local pre-processing – Image smoothing – Edge detectors – Zero-crossings of the second derivative – Scale in image processing – Canny edge detection – Parametric edge models – Edges in multi-spectral images – Local pre-processing in the frequency domain – Line detection by local preprocessing operators – Image restoration.
	UNIT III: **VGGNet and AlexNet Networks** Technical requirements , AlexNet and VGG Neural Networks, VGG16 architecture, Difference between VGG16 and VGG19, Developing solutions using AlexNet and VGG **Face Recognition And Gesture Recognition** Face Recognition-Introduction-Applications of Face Recognition-Process of

	Face Recognition DeepFace solution by Facebook-FaceNet for Face Recognition- Implementation using FaceNet-Gesture Recognition.
	UNIT IV: **VIDEO ANALYTICS** Video Processing – use cases of video analytics-Vanishing Gradient and exploding gradient problem-RestNet architecture-RestNet and skip connections-Inception Network-GoogleNet architecture-Improvement in Inception v2-Video analytics-RestNet and Inception v3.
Text books and Reference books	**Text Book(s):** [1]. Milan Sonka, Vaclav Hlavac, Roger Boyle, "Image Processing, Analysis, and Machine Vision", 4nd edition, Thomson Learning, 2013. [2]. VaibhavVerdhan,(2021, Computer Vision Using Deep Learning Neural Network Architectures with Python and Keras,Apress 2021 **Reference Books:** [1]. Richard Szeliski, "Computer Vision: Algorithms and Applications", Springer Verlag London Limited,2011. [2]. Caifeng Shan, FatihPorikli, Tao Xiang, Shaogang Gong, "Video Analytics for Business Intelligence", Springer, 2012. [3]. D. A. Forsyth, J. Ponce, "Computer Vision: A Modern Approach", Pearson Education, 2003. [4]. E. R. Davies, (2012), "Computer & Machine Vision", Fourth Edition, Academic Press.
E-resources and other digital material	[1]. Dr. Mubarak Shah, (19, 12, 2023). UCF Computer Vision Video Lectures,https://www.youtube.com/watch?v=715uLCHt4jE&list=PLd3hlSJsX_ImKP68wfKZJVIPTd8Ie5u-9 [2]. Dr. Andrew Ng, (19, 12, 2023).DeepLearningAI,https://www.youtube.com/@Deeplearningai/videos

Course Category:	Programme Elective – III	Credits:		3
Course Type:	Theory	Lecture-Tutorial-Practice:		3-0-0
Prerequisites:	-	Continuous Evaluation:		40
		Semester end Evaluation:		60
		Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:			
	CO1	Apply pre-processing techniques on text data.		
	CO2	Solve NLP problems using probabilistic language models		
	CO3	Analyze linguistic structure in text, using parsing and CFG		
	CO4	Construct syntactic and semantics structures for a given sentence		

Contribution of Course Outcomes towards achievement of Program Outcomes 1-Low, 2-Medium, 3-High)		PO1	PO2	PO 3	PO 4	PO 5
	CO1	1			2	1
	CO2	3			3	2
	CO3				1	3
	CO4			2	3	3

Course Content	**UNIT I:** **Regular Expressions, Text Normalization, Edit Distance:** Regular Expressions, Words, Corpora, Text Normalization, Word Tokenization, Word Normalization, Lemmatization and Stemming, Minimum Edit Distance, The Minimum Edit Distance Algorithm. **Natural Language Toolkit Essentials** (NLTK): **Introduction to NLP** – Diving into NLTK, **Text Wrangling and Cleansing** – What is Text wrangling? Text cleansing, Sentence splitter, Tokenization, Stemming, Lemmatization, Stop word removal, Rare word removal.
	UNIT II: **N-gram Language Models –** N Grams, Evaluating Language Models, Perplexity, Sampling sentences from a language model, Smoothing-Laplace smoothing, Add-k Smoothing, Back off and Interpolation. **Naive Bayes and Sentiment Classification –** Naive Bayes Classifiers, Training the Naïve Bayes Classifier, Worked example, Optimizing for Sentiment Analysis, Naive Bayes for other text classification tasks, Naive Bayes as a Language Model, Evaluation: Precision, Recall, F-measure.
	UNIT III: **Sequence Labeling for parts of Speech and Named Entities:** English Word Classes, Parts-of-speech-Tagging, Named Entities and Named Entities Tagging, HMM Parts of Speech Tagging, Markov Chains, The Hidden Markov Model, The components of an HMM tagger. **Context-Free Grammars and Constituency Parsing:**Constituency,

	Context-Free Grammars, Formal Definition of Context-Free Grammar, Treebanks, Grammar Equivalence and Normal Form, Ambiguity, CKY Parsing: A Dynamic Programming Approach - Conversion to Chomsky Normal Form, CKY Recognition, CKY Parsing, CKY in Practice. **Case Study:** Generation of Tags from a given sentences using Penn Treebank tagset.
	UNIT IV: **Logical Representations of Sentence Meaning:** Computational Desiderata for Representations, Model-Theoretic Semantics, First-Order Logic, Variables and Quantifiers, Lambda Notation, The Semantics of First-Order Logic, Inference, Event and State Representations, Description Logics. **Word Senses and WordNet:** Word Senses, Defining Word Senses, How many senses do words have, Relations Between Senses, WordNet: A Database of Lexical Relations, Sense Relations in WordNet. **Case Study:** Sentiment analysis of text data using NLTK
Text books and Reference books	**Text Book(s):** [1]. Daniel Jurafsky and James H.Martin,"Speech and Language Processing:An Introduction to Natural Language Processing, Computational Linguistics,andSpeechRecognition", (availbale at **https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf**) [2]. Jacob Perkins, Nitin Hardeniya, Deepti Chopra, ItiMathur, Nisheeth Josh Natural Language Processing: Python and NLTK, Packt Publishing,2016 **Reference Books:** [1]. Hobson lane, Cole Howard, Hannes Hapke, "Natural language processing in action" MANNING Publications, 2019. [2]. Rajesh Arumugam, RajalingappaShanmugamani "Hands-on natural language processing with python: A practical guide to applying deep learning architectures to your NLP application". PACKT publisher, 2018. [3]. Michael Nielsen,"Neural Networks and Deep Learning", Determination Press 2015
E-resources and other digital material	[1]. Dan Jurafsky and Christopher Manning, Natural Language Processig Course,Stanford,26thJun2019,https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html [2]. Dan Jurafsky and Christopher Manning, Natural Language Processig Course,10thJun2018,https://www.youtube.com/watch?v=3Dt_yh1mf_U&list=PLQiyVNMpDLKnZYBTUOlSI9mi9wAErFtFm [3]. Prof. DragomirRadev,Ph.D.,Lecture Series on Natural Language Processing,Department of School of Information,University of Michigan,Mar 2016 https://www.youtube.com/playlist?list=PLLssT5z_DsK8BdawOVCCaTCO99Ya58ryR [4]. https://www.nltk.org/howto.html

23ITDS2015A -WEB MINING AND SOCIAL NETWORK ANALYSIS

Course Category:	Programme Elective – IV		Credits:		4
Course Type:	Theory		Lecture-Tutorial-Practice:		4-0-0
Prerequisites:			Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100

Course Outcomes	Upon successful completion of the course, the student will be able to:	
	CO1	Apply graph basics to analyze the social media data and measure the network measures
	CO2	Derive the similarities of people in the society and find the communities in the society.
	CO3	Generate recommendations, social recommendations and evaluate recommendations.
	CO4	Measuring influence and homophily, Analyze the individual behavior and collective behavior

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	2	1	3	2	2
	CO2	3	1	3	2	2
	CO3	3	1	3	2	2
	CO4	2	1	3	2	2

Course Content	**UNIT I:** **Social media mining and its challenges. Graph Essentials:** Graph basics: nodes, edges, degree, degree distribution. Graph representation, types of graphs. Connectivity in graphs. Special graphs **Network measures**: centrality : degree centrality, eigenvector centrality, katz centrality, page rank, betweenness centrality
	UNIT II: **Community analysis**: Community Detection, Node degree, Node Reachability, Social Communities, Community Detection Algorithms, Member Based Community Detection **Group Based Community Detection Algorithms:** Balanced Communities, Robust Communities, Modular Communities, Dense Communities and Hierarchical communities.
	UNIT III: **Recommendation in Social Media:** Recommendation System challenges, classical recommendation algorithms, Content-based methods, collaborative filtering, User-based collaborative filtering, item-based collaborative filtering, **Recommendation using social context:** Recommendation using social context alone, recommendation constrained by social context, Evaluating

	Recommendations.
	UNIT IV: **Influence and Homophily:** Measuring assortativity, Measuring Assortativity for Nominal Attributes and for ordinal attributes. Influence: Measuring influence and modeling influence. Homophily: Measuring Homophily, modeling Homophily, Distinguish Influence and Homophily. Behavior Analytics: Individual behavior, Individual behavior analysis, behavior analysis methodology, individual behavior Modeling and prediction, link prediction
Text books and Reference books	**Text Book(s):** [1] Reza Zafarani, Mohammad Ali Abbasi , Huan Liu Social Media Mining: An Introduction **Reference Books:** [1]. Charu c. aggarwal "Social network data analytics" Springer [2] M. E. J. Newman Hardback "Networks: An Introduction by M. E. J. Newman, a college-level textbook about the science of networks.", Oxford University Press, 2010.
E-resources and other digital material	1. Dr Bernie Hoganhttps://www.youtube.com/watch?v=2zhuj8ubinmSocial network analysis - Introduction to structural thinking:, University of Oxford, 2018. 2. S.R.S. Lyengarhttps://www.youtube.com/watch?v=b7Ug1h6EGNk "Introduction to Social Networks, 2017.

23ITDS2015B-OPTIMIZATION TECHNIQUES FOR DATA ANALYSIS

Course Category:	Programme Elective-IV		Credits:		3
Course Type:	Theory		Lecture-Tutorial-Practice:		3-0-0
Prerequisites:			Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Summarize various techniques used for optimization problems arising from engineering areas.			
	CO2	Analyze optimization algorithms for Linear Programming problems			
	CO3	Solve various constrained and unconstrained nonlinear programming problems			
	CO4	Apply modern and multi objective optimization techniques to provide optimal solution for real time problems			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3-High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	3			3	1
	CO2	2			2	
	CO3	2			2	
	CO4	3			2	3

Course Content	**UNIT I:** **Introduction to Optimization:** Introduction, Historical Development, Engineering Applications of Optimization, Statement of an Optimization Problem, Classification of Optimization Problems. **Classical Optimization Techniques:** Single-Variable Optimization, Multivariable Optimization with No Constraints, Multivariable Optimization with Equality Constraints, Multivariable Optimization with Inequality Constraints.
	UNIT II: **Linear Programming** Introduction, Applications of Linear Programming, Standard Form of a Linear Programming Problem, Geometry of Linear Programming Problems, Solution of a System of Linear Simultaneous Equations, Pivotal Reduction of a General System of Equations **Simplex Method:**Motivation of the Simplex Method, Simplex Algorithm, Improving a Non-optimal Basic Feasible Solution, Two Phases of the Simplex Method and Applications of Simplex Algorithm.
	UNIT III: **Nonlinear Programming Algorithms**: Applications of Unconstrained Algorithms – Direct Search Method, Gradient method, Applications of Constrained Algorithms - Separable Programming, Quadratic Programming,

	Chance- Constrained Programming, Linear Combinations method, Applications of SUMT Algorithm. **Case Study 1**: Chance Constrained Problem
	UNIT IV: **Modern Methods of Optimization** Introduction, Applications of Genetic Algorithms, Applications of Simulated Annealing, Applications of Particle Swarm Optimization, Applications of Ant Colony Optimization, Optimization of Fuzzy Systems, Applications of Neural-Network-Based Optimization **Multi objective Optimization** Introduction, Pareto Solutions, Computing the Pareto Front, multi objective to single objective optimization. **Case Study 2:** Travelling Salesperson Problem
Text books and Reference books	**Text Book(s):** [1]. Singiresu S Rao, "Engineering Optimization Theory and Practice", John Wiley and sons, 4th Edition, 2009. [2]. HamdyA.Taha, "OprationResearch : An Introduction", 8th Edition, Pearson Prentice Hall, 2007. **Reference Books:** [1] S.Rao, "Engineering optimization: Theory and practice", 4th Edition, New Age International, 2009. [2] Edwin K. P. Chong and Stanislaw. Zak "An Introduction to Optimization", John Wiley and sons, 2nd Edition 2001. [3] Andreas Antoniou, "Practical Optimization Algorithms and Engineering Applications", [4] An Introduction to Optimization Edwin K., P. Chong & Stanislaw h. Zak.Andreas Antoniou. "Practical Optimization Algorithms and Engineering Applications", Springer Series, 2007. [5]. Paulo Cortez, "Modern Optimization with R", Springer series, 2014.
E-resources and other digital material	[1]. Prof.A.Goswami, Department of Mathematics, IIT Kharagpur, "Optimization", 2014 https://nptel.ac.in/courses/111105039/ [2].Dr.AdityaJagannadham, IIT Kanpur, "Applied Optimization for Wireless, Machine Learning and Big Data,2018, https://nptel.ac.in/courses/108104112/ [3]. Stephen J.Wright , "Optimization Algorithms for Data Analysis ", 2016 https://bfi.uchicago.edu/wp-content/uploads/StephenWrightPCMI2016FinalProofs.pdf [4].Geoff Gordon and Ryan Tibshirani,School of Computer Science, Carnegie Mellon University, " Optimization", 2012, https://www.cs.cmu.edu/~ggordon/10725-F12/schedule.html

23ITDS2015C-INFORMATION RETRIEVAL SYSTEMS

Course Category:	Programme Elective -IV	Credits:	3
Course Type:	Theory	Lecture-Tutorial-Practice:	3-0-0
Prerequisites:	Basic statistics	Continuous Evaluation:	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	Upon successful completion of the course, the student will be able to:	
	CO1	Interpret the basic concepts and techniques in Information Retrieval
	CO2	Evaluate information retrieval system performance and queries formulation
	CO3	Infer relevance feedback and query operations on a text database
	CO4	Analyze the web characterization, web search tasks. and digital libraries implications

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	`			1	1
	CO2	1				2
	CO3			2		2
	CO4	2			3	3

Course Content	**UNIT I:** **Boolean retrieval :** An example information retrieval problem , A first take at building an inverted index , Processing Boolean queries , The extended Boolean model versus ranked retrieval **The term vocabulary and postings lists**: Document delineation and character sequence decoding, Obtaining the character sequence in a document, Choosing a document unit, Determining the vocabulary of terms, Tokenization, Dropping common terms: stop words, Normalization (equivalence classing of terms) , Stemming and lemmatization
	UNIT II: **Scoring, term weighting and the vector space model :** Parametric and zone indexes , Weighted zone scoring , Learning weights , The optimal weight g , Term frequency and weighting , Inverse document frequency, Tf-idf weighting **The vector space model for scoring**: Dot products, Queries as vectors **Probabilistic information retrieval**: Review of basic probability theory, The Probability Ranking Principle, The 1/0 loss case The PRP with retrieval costs
	UNIT III: **Text classification and Naive Bayes**:The text classification problem, Naive Bayes text classification, Relation to multinomial unigram language model, The Bernoulli model, Properties of Naive Bayes, A variant of the multinomial model **Web search basics** :Background and history ,Web characteristics, The web graph, Spam, Advertising as the economic model, The search user experience ,

	User query needs
	UNIT IV: **Web crawling and indexes** :Overview, Features a crawler must provide, Features a crawler should provide, Crawling, Crawler architecture,DNS resolution, The URL frontier Link analysis : The Web as a graph, Anchor text and the web graph, Page Rank , Markov chains, The Page Rank computation Case study : Implementation of various classification algorithms on text, Design and development of Question/Answering System
Text books and Reference books	**Text Book(s):** [1].Christopher D. Manning, PrabhakarRaghavan and HinrichSchütze, Introduction to Information Retrieval, Cambridge University Press. 2008. [2].ChengXiangZhai, Statistical Language Models for Information Retrieval (Synthesis Lectures Series on Human Language Technologies), Morgan & Claypool Publishers, 2008 **Reference Books:** **[1].**G. G. Chowdhurry,Introduction to Modern Information Retrieval, Neal-Schuman Publishers; Third edition , 2019 **[2].**Gerald J Kowalski, Mark T Maybury Information Storage and Retrieval Systems: Theory and Implementation, Springer, 2004.
E-resources and other digital material	[1].Informationretrieval,Prof.Pabitramitra,IITkharagpur,http://cse.iitkgp.ac.in/~pabitra/course/ir06/ir06.html [2].Informationretrieval,Prof.Pawangoyal,IITkharagpur,http://cse.iitkgp.ac.in/~pawang/courses/ir16/lec1.html [3].Natural language processing by prof. Pushpakbhattacharyya, Department ofComputer science &engineering, IIT bombay,Https://www.youtube.com/watch?v=m0oiaogsqfw [4].Introduction to information retrieval, university of south carolina, https://www.youtube.com/watch?v=yluvahnq3wk

23ITDS2051-BIG DATA LAB

Course Category:	Laboratory - I	Credits:	1.5
Course Type:	Lab	Lecture-Tutorial-Practice:	0-0-3
Prerequisites:	Database Management Systems Lab	Continuous Evaluation:	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	Upon successful completion of the course, the student will be able to:	
	CO1	Implement Hdfs And Map Reduce Paradigm For Batch Oriented Applications.
	CO2	Apply Nosql Concepts To Store And Process Varieties Of Data.
	CO3	Solve Data Intensive Problems Using Pig Latin And Hive.
	CO4	Develop Solutions For Real Time Problems Using Spark.

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	2		1		3
	CO2	3		1		3
	CO3	3		1		3
	CO4	3		1		3

Course Content	**WEEK 1:**
	• Introduction, Applications, Tools related to big data and NOSQL.
	WEEK 2:
	• NOSQL: Mangodb installation and querying in Mangodb.
	WEEK 3: **Querying in Mangodb**
	• Create Database
	• Drop Database
	• Create collection
	• Drop collection
	• Indexing
	• Aggregation
	WEEK 4:
	• Installation Of Cloudera
	WEEK 5:
	• Exploring HDFS and Listing of files.
	WEEK 6: HDFS Operations using various commands.
	WEEK 7: **HiveQL**

	Create Database
	Drop Database
	Create table
	Alter table
	Drop table
	Partitioning
	Built-in operators
	Built-in functions
	Views and indexes

	WEEK 8: **HiveQl** • Select where • Order by • Group by • Joins
	WEEK 9: **Map Reduce Applications** • Mapper code • Reducer code • Combiner code
	WEEK 10: **Pig Latin Scripts** • Operators • Load & Store • Diagnostic • Grouping and Joining • Combining and Splitting • Filtering • Sorting
	WEEK 11: **Spark SQL**
	WEEK 12: **Case Study on Hive and Pig from kaggle**
	WEEK 13: **Case Study on Map reduce**
Text books and Reference books	**Text Book(s):** [1]. Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, "Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data" 1st Edition, TMH,2012. [2]. Tom White, Hadoop, "The Definitive Guide", 3rd Edition, O'Reilly Publications, 2012. **Reference Books:** [1]. SeemaAcharya, SubhashiniChellappan, Big Data and Analytics, Wiley Publishers. [2]. Holden Karau, Andy Konwinski, Patrick Wendell, MateiZaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly Media, Inc.

E-resources and other digital material	[3]. **Pig Tutorial for Practice, https://www.tutorialspoint.com/apache_pig/index.htm**
	[4]. Mangodb Tutorial For Practice, Pig Tutorial For Practice, https://www.tutorialspoint.com/Mongodb/Index.htm [5]. Hive Tutorial For Practice, https://www.tutorialspoint.com/Hive/Index.htm

23ITDS2052-DATA VISUALIZATION LAB

Course Category:	Laboratory-II		Credits:		1.5
Course Type:	Lab		Lecture-Tutorial-Practice:		0-0-3
Prerequisites:	Any programming language		Continuous Evaluation:		40
			Semester end Evaluation:		60
			Total Marks:		100
Course Outcomes	Upon successful completion of the course, the student will be able to:				
	CO1	Understand the visualization pipeline with its relationship to other data			
	CO2	Design considerations for the components of the good visualization			
	CO3	Construct visualizations for different attributes and showcase them in plots, interpret using R/Python			
	CO4	Construct visualizations for effective data analysis			

Contribution of Course Outcomes towards achievement of Program Outcomes (1-Low, 2-Medium, 3- High)		PO1	PO2	PO3	PO 4	PO 5
	CO1	`			1	1
	CO2	1				2
	CO3			2		2
	CO4	2			3	3

Course Content	Week 1: Implement Acquiring and plotting data like Pie chart, Area Chart and Bubble plot on real-time data
	Week 2: Implement Statistical Analysis – such as Multivariate Analysis, PCA, LDA, Correlation regression and analysis of variance.
	Week 3 & 4: Implementing data visualization using R 1. Find the data distributions using box and scatter plot. 2. Find the outliers using plot. 3. Plot the histogram, bar chart and pie chart on sample data.
	Week 5 & 6: Implementing basic operations in Tableau to get accustomed to its interface and Emphasizing the Results and Map View [1] Tableau Workspace, Connecting to a Data Source, Creating a view and Refining the view [2] Adding Filters to the view, Adding Colors to the view and Key Findings [3] Building a Map View, Getting into details and Identifying the Key Points
	Week 7: Creating a dashboard and building story to showcase stories in presentation mode [1] Creating a dashboard and Adding Instructiveness [2] Building a Story and Making a Conclusion [3] Visualization on Streaming dataset (Stock market dataset, weather forecasting).
	Week 8: Tracking Twitter data to see how fast information spreads online: Create a data visualization to understand the spread of information and miss information insights of individual tweets online.
	Week 9: Text visualization using web analytics

	Week 10: Motivate sales teams by modelling commission rates: Create a visualization to explore the relationships between compensation type, commission for sales people to motivate them.
	Week 10: Design a dashboard for real-time application of various massive dataset - Finance - Healthcare - Census –Geospatial.
Text books and Reference books	**Text Book(s):** [1]. Matthew Ward, Georges Grinstein and Daniel Keim, ―Interactive Data Visualization Foundations, Techniques, Applications‖,2010. [2]. Colin Ware, ―Information Visualization Perception for Design‖, 2nd edition, Margon Kaufmann Publishers,2004. **Reference Books:** [1]. Robert Spence ―Information visualization – Design for interaction‖, Pearson Education, 2 nd Edition,2007. [2]. . Alexandru C. Telea, ―Data Visualization: Principles and Practice,‖ A. K. Peters Ltd,2008.
E-resources and other digital material	[1]. Dr. Gauravdixit,department of management studies, indian institute of technology,Roorkee: https://nptel.ac.in/courses/110107092/7,2017 [2]. Padammarcus, and eugenewu. Res.6-009 how to process, analyze and Visualize data. January iap 2012. Massachusetts institute of technology: mit open Courseware, https://ocw.mit.edu.,2012 [3]. Prof.shankarnarasimhan,ragunathan, rengasamy,iit madras data Visualizationin r basicgraphics,Https://nptel.ac.in/courses/106106179/11,2016 [4]. Statistics and visualization for data analysis and inference, dr. Ed vul, Dr. Mike frank, massachusetts institute of technology,https://ocw.mit.edu/resources/res-9-0002-statistics-and-visualization-fordata- analysis-and-inference-january-iap-2009/, 2009.