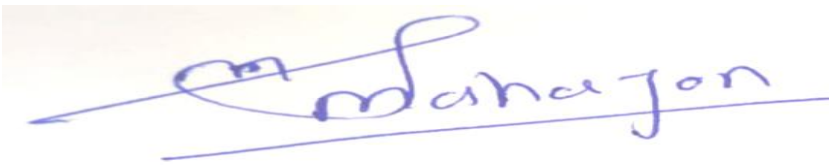
INDUS INSTITUTE OF TECHNOLOGY & ENGINEERING

Department of Computer Engineering

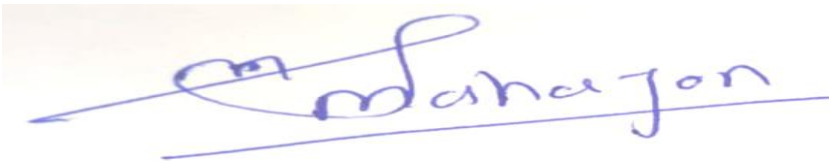
Proposed Teaching scheme and Syllabus
M. Tech. in Data Science & Analysis

INDUS UNIVERSITY
INDUS INSTITUTE OF TECHNOLOGY AND ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

M-TECH COMPUTER SCIENCE & ENGINEERING (Data Science & Analysis) SEMESTER –I TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019												
SR NO	**CODE**	**SUBJECTS**	**TEACHING SCHEME**			**CREDITS**	**HOURS**	**EXAMINATION SCHEME**				
								THEORY		**PRACT**		**TOTAL**
			L	**T**	**P**			**CIE**	**ESE**	**CIE**	**ESE**	
1	DS0101	Distributed Operating System	3	0	2	4	5	60	40	60	40	200
2	DS0102	Advanced Data Structures	3	0	2	4	5	60	40	60	40	200
3	DS0103	Data Science	2	0	2	3	4	60	40	-	-	100
4	DS0110	Research Methodology	3	0	0	3	3	60	40	-	-	100
5	DS0104	Optimization Technique (Elective – I)	3	0	2	4	5	60	40	60	40	200
	DS0105	Distributed Database (Elective – I)										
	DS0106	Data Preparation and Analysis (Elective – I)										
6	DS0107	Recommender System (Elective – II)	3	0	2	4	5	60	40	60	40	200
	DS0108	Machine Learning (Elective – II)										
	DS0109	Data Storage Technologies and Networks (Elective – II)										
TOTAL			**17**	**0**	**10**	**22**	**27**	**360**	**240**	**240**	**160**	**1000**

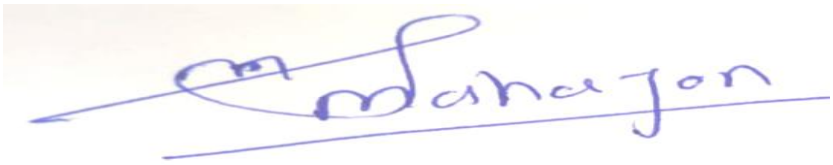


M-TECH COMPUTER SCIENCE & ENGINEERING (Data Science & Analysis) SEMESTER –II TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019												
SR NO	CODE	SUBJECTS	TEACHING SCHEME			CREDITS	HOURS	EXAMINATION SCHEME				
								THEORY		PRACT		TOTAL
			L	T	P			CIE	ESE	CIE	ESE	
1	DS0201	Advance Algorithms	3	0	2	4	5	60	40	60	40	200
2	DS0202	Soft Computing	3	0	2	4	5	60	40	60	40	200
3	DS0203	Cloud Computing	3	0	2	4	5	60	40	60	40	200
4	DS0204	Personality Enhancement	0	0	4	2	2	60	40	60	40	100
5	DS0205	Data Visualization (Elective – III)	3	0	2	4	5	60	40	60	40	200
	DS0206	Knowledge Discovery (Elective – III)										
	DS0207	Data Warehouse and Data Mining (Elective – III)										
6	DS0208	Data Security and Access Control (Elective – IV)	3	0	2	4	5	60	40	60	40	200
	DS0209	Web Analytics and Development (Elective – IV)										
	DS0210	Big Data Analytics (Elective – IV)										
TOTAL			**17**	**0**	**10**	**22**	**27**	**360**	**240**	**300**	**200**	**1100**



M-TECH COMPUTER SCIENCE & ENGINEERING (DATA SCIENCE & ANALYSIS) SEMESTER –III TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019												
SR NO	CODE	SUBJECTS	TEACHING SCHEME			CREDITS	HOURS	EXAMINATION SCHEME				
								THEORY		PRACT		TOTAL
			L	T	P			CIE	ESE	CIE	ESE	
1	DS0301	Dissertation Phase - I	0	0	24	12	24	-	-	60	40	100
TOTAL			0	0	24	12	24	-	-	60	40	100

M-TECH COMPUTER SCIENCE & ENGINEERING (DATA SCIENCE & ANALYSIS) SEMESTER –IV TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019												
SR NO	CODE	SUBJECTS	TEACHING SCHEME			CREDITS	HOURS	EXAMINATION SCHEME				
								THEORY		PRACT		TOTAL
			L	T	P			CIE	ESE	CIE	ESE	
1	DS0401	Dissertation Phase - II	0	0	24	12	24	-	-	60	40	100
TOTAL			0	0	24	12	24	-	-	60	40	100



SEMESTER-1

M-TECH COMPUTER SCIENCE & ENGINEERING (Data Science & Analysis) SEMESTER –I TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019												
SR NO	CODE	SUBJECTS	TEACHING SCHEME			CREDITS	HOURS	EXAMINATION SCHEME				
								THEORY		PRACT		TOTAL
			L	T	P			CIE	ESE	CIE	ESE	
1	DS0101	Distributed Operating System	3	0	2	4	5	60	40	60	40	200
2	DS0102	Advanced Data Structures	3	0	2	4	5	60	40	60	40	200
3	DS0103	Data Science	2	0	2	3	4	60	40	-	-	100
4		Research Methodology	3	0	0	3	3	60	40	-	-	100
5	DS0104	Optimization Technique (Elective – I)	3	0	2	4	5	60	40	60	40	200
	DS0105	Distributed Database (Elective – I)										
	DS0106	Data Preparation and Analysis (Elective – I)										
6	DS0107	Recommender System (Elective – II)	3	0	2	4	5	60	40	60	40	200
	DS0108	Machine Learning (Elective – II)										
	DS0109	Data Storage Technologies and Networks (Elective – II)										
TOTAL			**17**	**0**	**10**	**22**	**27**	**360**	**240**	**240**	**160**	**1000**

Subject: **Distributed Operating System**								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0101			Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	5	40	40	60	60	200

Course Objectives:

1. List the principles of distributed systems and describe the problems and challenges associated with these principles.
2. Understand Distributed Computing techniques, Synchronous and Processes.
3. Apply Shared Data access and Files concepts.
4. Design a distributed system that fulfills requirements with regards to key distributed systems
5. properties.
6. Understand Distributed File Systems and Distributed Shared Memory.
7. Apply Distributed web-based system.
8. Understand the importance of security in distributed systems

Course Content

UNIT-I

[12 hours]

Introduction to distributed Systems:
Definition and goals, Hardware and Software concepts, Design issues

Communication in Distributed System:
Computer Network and Layered protocols, Message passing and related issues, synchronization, Client Server model & its implementation, remote procedure call and implementation issues, Case Studies: SUN RPC, DEC RPC

UNIT-II

[12 hours]

Synchronization in distributed systems:
Clock synchronization and related algorithms, mutual exclusion, Deadlock in distributed systems

Processes and processors in distributed systems:
Threads, system model, processor allocation, scheduling in distributed systems: Load balancing and sharing approach, fault tolerance, Real time distributed systems, Process migration and related issues.

UNIT-III

[12 hours]

Distributed File Systems:
Introduction, features & goal of distributed file system, file models, file accessing models, file sharing semantics, file caching scheme, file replication, fault tolerance, trends in distributed file system

Distributed Shared Memory:
Introduction, general architecture of DSM systems, design and implementation issues of DSM, granularity, structure of shared memory space, consistency models, replacement strategy, thrashing

UNIT-IV

[12 hours]

Naming
Overview, Features, Basic concepts, System oriented names, Object locating mechanisms, Issues in designing human oriented names, Name caches, Naming and security, DNS

Distributed Web based Systems
Architecture, Processes, Communication, Naming, Synchronization, Consistency and Replication: Web Proxy Caching, Replication for Web Hosting Systems, Replication of Web Applications

Security
Introduction of Security in Distributed OS, Overview of security techniques, features, Need, Access Control, Security Management

Course Outcomes:

The students should be able to:

1. List the principles of distributed systems and describe the problems and challenges associated with these principles.
2. Understand Distributed Computing techniques, Synchronous and Processes.
3. Apply Shared Data access and Files concepts.
4. Design a distributed system that fulfills requirements with regards to key distributed systems properties.
5. Understand Distributed File Systems and Distributed Shared Memory.
6. Apply Distributed web-based system.
7. Understand the importance of security in distributed systems

TEXT BOOKS

1. Distributed Operating Systems Concepts and Design, Pradeep K. Sinha, PHI
2. Distributed Systems: Concepts and Design by George Coulouris, Jean Dollimore , Tim Kindberg, Pearson

REFERENCES

1. Distributed Operating Systems by Andrew S Tannebaum, Pearson
2. Distributed Computing by Sunita Mahajan & Seema Shah OXFORD
3. Distributed Systems: Principles and Paradigms by Andrew S Tanebaum, Maarten Van Steen, PHI
4. Distributed Computing, Fundamentals, Simulations and Advanced topics, 2nd Edition, Hagit Attiya and Jennifer Welch, Wiley India

WEB RESOURCES

https://www.dos.org/

https://www.tutorialpoint.com/dos/

LIST OF EXPERIMENTS

Exp No.	Title	Learning Outcomes
1.1	Write a Program to implement Concurrent Echo Client Server Application.	Basic Knowledge of client server model.
2.1	Write the Programs for Remote Procedure call.	Basic Knowledge of RPC.
3.1	Write the Programs for Remote Method Invocation.	Basic Knowledge of RMI.
4.1	Write the Programs for Thread Programming in JAVA	Basic Knowledge of Thread
4.2	Implement CORBA file.	Basic of CORBA
5.1	Write a Program to Increment a Counter in Shared Memory..	Basic Knowledge of Shared Memory
5.2	Implement Network File System (NFS).	Basic Knowledge of Shared Memory
5.3	Creation of a BPEL (Business Process Execution Language) Module and a Composite Application.	Basic Knowledge of Shared Memory
5.4	Study of Grid Services using various Tools.	Basic Knowledge of Shared Memory

Analysis								
Teaching Scheme				Examination Evaluation Scheme				
Lecture	Tutorial	Practical	Credits	University Theory Examination	University Practical Examination	Continuous Internal Evaluation (CIE)- Theory	Continuous Internal Evaluation (CIE)- Practical	Total
3	0	2	5	40	40	60	60	200

COURSE OBJECTIVE

1. The student should be able to choose appropriate data structures, understand the ADT/libraries, and use it to design algorithms for a specific problem.
2. Students should be able to understand the necessary mathematical abstraction to solve problems.
3. To familiarize students with advanced paradigms and data structure used to solve algorithmic problems.
4. Student should be able to come up with analysis of efficiency and proofs of correctness.

Course Content

Unit I

[07 hours]

Dictionaries: Definition, Dictionary Abstract Data Type, Implementation of Dictionaries. **Hashing:** Review of Hashing, Hash Function, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing, Linear Probing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing.

Unit II

[14 hours]

Skip Lists: Need for Randomizing Data Structures and Algorithms, Search and Update Operations on Skip Lists, Probabilistic Analysis of Skip Lists, Deterministic Skip Lists
Trees: Binary Search Trees, AVL Trees, Red Black Trees, 2-3 Trees, B-Trees, Splay Trees.

Unit III

[12 hours]

Text Processing: Sting Operations, Brute-Force Pattern Matching, The Boyer-Moore Algorithm, The Knuth-Morris-Pratt Algorithm, Standard Tries, Compressed Tries, Suffix Tries, The Huffman Coding Algorithm, The Longest Common Subsequence Problem (LCS), Applying Dynamic Programming to the LCS Problem.

Unit IV

[15 hours]

Computational Geometry: One Dimensional Range Searching, Two Dimensional Range Searching, Constructing a Priority Search Tree, Searching a Priority Search Tree, Priority Range Trees, Quadtrees, k-D Trees.

Recent Trends in Hashing, Trees, and various computational geometry methods for efficiently solving the new evolving problem

COURSE OUTCOMES

After completion of course, students would be able to:

1. Understand the implementation of symbol table using hashing techniques.
2. Develop and analyze algorithms for red-black trees, B-trees and Splay trees.
3. Develop algorithms for text processing applications.
4. Identify suitable data structures and develop algorithms for computational geometry problems.

Text Book:

1. Mark Allen Weiss, Data Structures and Algorithm Analysis in C++, 2nd Edition, Pearson, 2004.

References:

2. M T Goodrich, Roberto Tamassia, Algorithm Design, John Wiley, 2002.
3. Data Structures and Algorithms in C by Mark Allen Weiss - Pearson
4. The Design and analysis of computer Algorithms by A. V. Aho, R.Sethi and J. D. Ullman, Pearson.
5. Randomized Algorithms by Rajeev Motwani, Prabhakar Raghavan-Cambridge Univ.

LIST OF EXPERIMENTS

Sr No	Aim	Learning outcome
1	WAP for AVL Tree to implement following operations: (For nodes as integers)	Implement the height balancing concept of

	a. Insertion: Test program for all cases (LL, RR, RL, LR rotation)	tree using different methods.
	b. Deletion: Test Program for all cases (R0, R1, R-1, L0, L1, L-1)	
	c. Display: using set notation.	
2	Implement the above program I for nodes as Student structure, with key as Student_ roll_no.	
3	WAP to implement Red-Black trees with insertion and deletion operation for the given input data as Strings	
4	WAP to implement insertion, deletion, display and search operation in m-way B tree (i.e. a non-leaf node can have atmost m children) for the given data as integers (Test the program for m=3, 5, 7).	Implement the concept to be used in file storage structure.
5	WAP to implement insertion, deletion and display operation in Min-Max Heap for the given data as integers.	To implement data structure which will be used for implementing advanced algorithms.
6	WAP to implement Make_Heap, Insertion, Find_Min, Extract_Min, Union, Decrease_Key and Delete_Key operations in Binomial Heap for the given data as strings.	
7	WAP to implement priority queue.	
8	WAP to implement Make_Set, Find_Set and Union functions for Disjoint Set Data Structure for a given undirected graph G(V,E) using the linked list representation with simple implementation of Union operation.	
9	WAP to implement Dijkstra's algorithm for single-source shortest path in a weighted directed graph using fibonacci heap.	Implement and understand shortest path Data structure that is used in networking applications.
	WAP to store k keys into an array of size n at the location computed using a hash function, loc = key % n, where k<=n and k takes values from [1 to m], m>n. To handle the collisions use the following collision resolution techniques,	Implement the hashing concept that is used in various database application.
	a. Linear probing	
	b. Quadratic probing	
	c. Random probing	
	d. Double hashing/rehashing	
	e. Chaining	

Subject: Data Science			
Program: **M.Tech. in Data Science & Analysis**	Subject Code: DS0103	Semester**: I**	
Teaching Scheme	**Examination Evaluation Scheme**		

Lecture	Tutorial	Practical	Credits	University Theory Examination	University Practical Examination	Continuous Internal Evaluation (CIE)- Theory	Continuous Internal Evaluation (CIE)- Practical	Total
3	0	2	4	40	40	60	60	200

Course objectives:

1. Learn the fundamentals of data analytics and the data science pipeline
2. Learn how to scope the resources required for a data science project
3. Apply principles of Data Science to the analysis of business problems.
4. Apply data mining software to solve real-world problems.
5. Employ cutting edge tools and technologies to analyze Big Data

Course Content

UNIT-I

[12 hours]

Introduction

Defining Data Science, what do data science people do? Current landscape of perspectives, Data Science in Business, Use Cases for Data Science

UNIT-II

[12 hours]

Statistical Inference

Statistical modeling, probability distributions, fitting a model, Intro to R

Descriptive Statistics

Introduction to the course, Descriptive Statistics, Probability Distributions

Inferential Statistics

Inferential Statistics through hypothesis tests, Permutation & Randomization Test

UNIT-III

[12 hours]

Machine Learning Introduction and Concepts

Differentiating algorithmic and model based frameworks, Regression: Ordinary Least Squares, Ridge Regression, Lasso Regression, K Nearest Neighbors, Regression & Classification

Supervised Learning with Regression and Classification techniques

Bias-Variance Dichotomy, Model Validation Approaches, Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Regression and

Classification Trees, Support Vector Machines, Ensemble Methods: Random Forest, Neural Networks, Deep learning

Unsupervised Learning and data modelling

Clustering, Associative Rule Mining, Logical Modelling: Converting a conceptual model to logical model , Integrity constraints, Normalization.

UNIT-IV

[12 hours]

Data Visualization

Basic principles, ideas and tools for data visualization

Data Science and Ethical Issues

Discussions on privacy, security, ethics

Course Outcome:

- Students will demonstrate knowledge of big data analytics.
- Students will demonstrate the ability to think critically in making decisions based on data
- Students will able to interpret data, extract meaningful information, and assess findings.
- Students will identify and analyze social, legal, and ethical issues in data science.
- Students will be able to choose and apply tools and methodologies to solve data science tasks.

TEXT BOOKS

1. Applied statistics and probability for engineers, Montgomery, Douglas C., George C. Runger, John Wiley & Sons, 2010
2. Mining of Massive Datasets , Anand Rajaraman and Jeffrey David Ullman 2012
3. Doing Data Science, Straight Talk From The Frontline., Cathy O'Neil and Rachel Schutt , O'Reilly. 2014

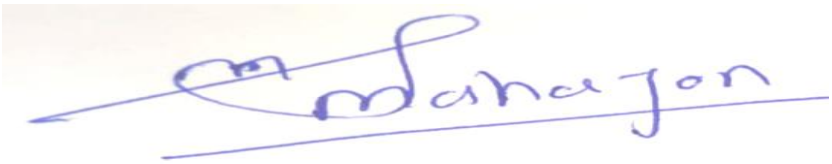
REFERENCES

1. Machine Learning: A Probabilistic Perspective. Kevin P. Murphy.

LIST OF EXPERIMENTS

Experiment. No.	Title
1	Getting Started with Skills Network Labs.

2	LAB: Getting Started with Jupyter Notebooks
3	LAB: Getting Started with Apache Zeppelin Notebooks
4	LAB: Getting Started with RStudio IDE
5	Data Analysis with Python Import data sets Clean and prepare data for analysis Manipulate pandas DataFrame Summarize data Build machine learning models using scikit-learn Build data pipelines
6	Data Visualization with Python Introduction to Visualization Tools Basic Visualization Tools Specialized Visualization Tools Creating Maps and Visualizing Geospatial Data
7	Advanced Visualization Tools



Subject: Optimization Technique								
Program: **M.Tech. in Data Science & Analysis**			Subject Code: DS0104				Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	0	3	40		60		100

Course Objective:
The aim of this subject is to provide a student with a broad and in depth knowledge of a range of operation research models and techniques, which can be applied to a variety of industrial applications.

Content

Unit I

(10 Hours)

Mathematical Preliminaries and basics of linear programming

- Open and Closed sets in Euclidean space, convex linear combinations, convex sets
- Intersection of convex sets, convex hull of a set
- Vertices or extreme points of a convex set, convex Polyhedron
- Introduction to linear programming, LP in two dimensional space, general LP problem
- Feasible solutions, basic solutions, basic feasible solutions, optimal Solutions
- Graphical method

Unit II:

(13 Hours)

Problem solving techniques for LP problems

- Simplex method
- Canonical form of equations
- Finding the first basic feasible solutions, artificial variables
- Degeneracy
- Simplex Multipliers
- Revised simplex methods
- Duality in LP problems
- Application of Duality

Unit III:

(12 Hours)

Transportation and Assignment Problems

- Introduction of transportation problems
- Transportation array, transportation matrix, triangular basis
- Finding a basic feasible solution
- Loop in transportation array
- Assignment problems: Formulation, Hungarian Method, Constrained Assignment Problem.

Unit IV

(10 Hours)

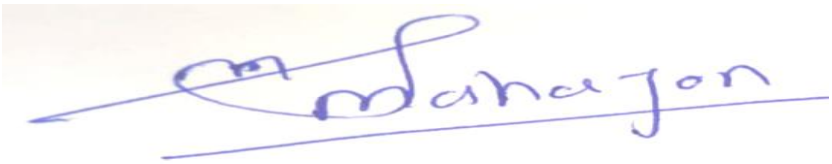
Theory of games and Sequencing problems

- Study of Two Person Zero-sum Game Problems,
- Games with and without Saddle point,
- Principles of Dominance, Graphical method,
- Conversion of Game problem into an LP problem
- Problem of Sequencing: Sequencing of n jobs on two and three machines,
- Applications of Game theory and Sequencing problems

Text Book: Optimization methods in operation research and system analysis, by K V Mittal, C Mohan, 3rd edition, New Age International (P) Limited, Publishers, 1996

Reference books:

1. Operation Research: Theory and Applications by J K Sharma, 6th edition, Trinity press
2. Operation Research - S. D. Sharma, Kedarnath Ramnath & Co.
3. Operation Research - Kanti Swaroop & Man Mohan, Sultan Chand & Co.
4. Operation Research – Hamdy A. Taha, Pearson Education
5. Linear Programming - L. I. Gass, Tata MacGraw Hills book-company.



Subject: Research Methodology and IPR								
Program: **M.Tech. in Data Science & Analysis**			Subject Code:				Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	0	3	40	-	60	-	100

Course objectives:

1. The main objective of the IPR is to make the students aware of their rights for the protection of their invention done in their project work.

2. To get registration in our country and foreign countries of their invention, designs and thesis or theory written by the students during their project work and for this they must have knowledge of patents, copy right, trademarks, designs and information Technology Act.
3. Further teacher will have to demonstrate with products and ask the student to identify the different types of IPR's.

Course Content

UNIT-1

12

RESEARCH FORMULATION AND DESIGN

Motivation and objectives –Research methods vs. Methodology. Types of research – Descriptive vs Analytical, applied vs Fundamental, Quantitative Vs Qualitative, Conceptual vs Empirical, concept of applied and basic research process, criteria of good research. Defining and formulating the research problem, selecting the problem, necessity of defining the problem, importance of literature review in defining a problem, literature review primary and secondary sources, reviews, monograph, patents, research databases, web as a source, searching the web, critical literature review, identifying gap areas from literature and research database, development of working hypothesis

UNIT-2

12

DATA COLLECTION AND ANALYSIS

Accepts of method validation, observation and collection of data, methods of data collection, sampling methods, data processing and analysis strategies and tools, data analysis with statically package (Sigma STAT, SPSS for student t-test, ANOVA, etc.), hypothesis testing.

SOFT COMPUTING

Computer and its role in research, Use of statistical software SPSS, GRETL etc. in research. Introduction to evolutionary algorithms Fundamentals of Genetic algorithms, Simulated Annealing, Neural Network based optimization, Optimization of fuzzy systems.

UNIT-3

12

RESEARCH ETHICS, IPR AND SCHOLARY PUBLISHING

Ethics-ethical issues, ethical committees (human & animal); IPR-intellectual property rights and patent law, commercialization, copy right, royalty, trade related aspects of intellectual property rights (TRIPS); scholarly publishing-IMRAD concept and design of research paper, citation and acknowledgement, plagiarism, reproducibility and accountability.

UNIT-4

08

INTERPRETATION AND REPORT WRITING

Meaning of Interpretation, Technique of Interpretation, Precaution in Interpretation, Significance of Report Writing, Different Steps in Writing Report, Layout of the Research

Report Types of Reports, Oral Presentation, Mechanics of Writing a Research Report, Precautions for Writing Research Reports, Conclusions.

Course Outcomes:

At the end of the course, students will demonstrate their ability to:

1. Understanding and formulation of research problem.
2. Analyze research related information.
3. Understand plagiarism and follow research ethics
4. Understand that today's world is controlled by Computer, Information Technology, but tomorrow world will be ruled by ideas, concept, and creativity.

TEXT BOOKS

1. Garg, B.L., Karadia, R., Agarwal, F. and Agarwal, U.K., 2002. An introduction to Research Methodology RBSA Publishers
2. Kothari, C.R., 1990. Research Methodology: Methods and Techniques New Age International. 418p.

REFERENCES

1. Sinha, S.C. and Dhiman, A.K., 2002. Research Methodology Ess Ess Publications.2 volumes.
2. Trochim, W.M.K., 2005. Research Methods: the concise knowledge base Atomic Dog Publishing. 270p.
3. Wadehra, B.L. 2000. Law relating to patents, trademarks, copyright designs and geographical indications Universal Law Publishing

Subject: Distributed Database (Department Elective - I)								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0105			Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

Course objectives:

The objective of course is to provide insight to distributed database, normalization techniques and integrity rules. It also includes parallel database systems along with object oriented models

Course Content

UNIT-I [12 hours]

Introduction: Distributed Data processing, Distributed database system (DDBMS), Promises of DDBMSs, Complicating factors and Problem areas in DDBMSs, Overview Of Relational DBMS Relational Database concepts, Normalization, Integrity rules, Relational Data Languages, Relational DBMS.

UNIT-II [12 hours]

Distributed DBMS Architecture: DBMS Standardization, Architectural models for Distributed DBMS, Distributed DBMS Architecture. Distributed Database Design: Alternative design Strategies, Distribution design issues, Fragmentation, Allocation. Semantic Data Control: View Management, Data security, Semantic Integrity Control.

UNIT-III [12 hours]

Overview of Query Processing: Query processing problem, Objectives of Query Processing, Complexity of Relational Algebra operations, characterization of Query Processors, Layers of Query Processing. Introduction to Transaction Management: Definition of Transaction, Properties of transaction, types of transaction. Distributed Concurrency Control: Serializability theory, Taxonomy of concurrency control mechanisms, locking bases concurrency control algorithms

UNIT-IV **[12 hours]**

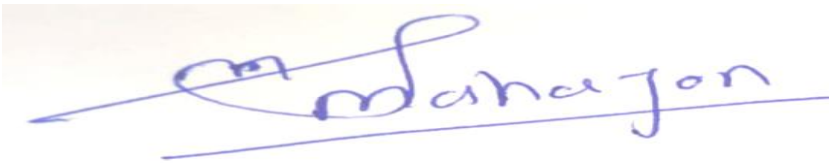
Distributed Object Database Management systems: Fundamental Object concepts and Object models, Object distribution design. Architectural issues, Object management, Distributed object storage, Object query processing. Transaction
Management. Database Interoperability: Database Integration, Query processing

Course Outcome:

Able to understand relational database management systems, normalization to make efficient retrieval from database and query

REFERENCES

1. Principles of Distributed Database Systems, Second Edition, M. Tamer Ozsu Patrick Valduriez
2. Distributed Databases principles and systems, Stefano Ceri, Giuseppe Pelagatti, Tata McGraw Hill.



Subject: Data Preparation and Analysis (Department Elective - I)								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0106			Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

Course objectives:

1. To learn how to gather and analyze large sets of data to gain useful business understanding and how to produce a quantitative analysis report/memo with the necessary information to make Decisions.
2. To survey industrial and scientific applications of Data Analytics, with case studies.
3. To prepare data for analytics and perform exploratory data analysis.
4. To develop meaningful data visualizations.
5. To perform cleaning and reformatting real world data for analysis.
6. To apply summary statistics techniques over datasets.

Course Content

UNIT-I

Introduction to Data Analysis **[10 hours]**

Defining data analysis problems: Knowing the client, Understanding the questions; Data Gathering and Preparation: Data formats, parsing and transformation, Scalability and real-time issues

UNIT-II

Exploratory Analysis **[12 hours]**

Data Cleaning: Consistency Checking, Heterogeneous and Missing data, Data Transformation and Segmentation; Exploratory Analysis: Descriptive and Comparative Analysis, Clustering and Association, Hypothesis Generation

UNIT-III

Visualization **[08 hours]**

Designing Visualizations, Time Series, Geolocated Data, Correlations and Connections, Hierarchies and Networks, Interactivity

UNIT-IV

Ethics in the Profession **[06 hours]**

Cases in Computing, Statistics and Communication, Professional ethics codes: ACM, IEEE, AM Stat. Assoc.

Course Outcome:

At the end of the course, students will demonstrate the ability to

1. Apply Clean and Format Real Time Data pertaining to Real Time Data Science Applications
2. Visualize Data in Multiple Dimensions as per the application requirement
3. Draw a comparative and descriptive analysis of the data
4. Obtain results by applying statistic techniques over datasets

TEXT BOOKS

1. Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt

REFERENCES

1. Data Analytics by Anil Maheshwari – McGraw Hill
2. Data Preparation for Data Mining by Dorian Pyle – Morgan Kaufmann Series
3. Data Preparation in the Big Data by Federico Castanedo – O'Reilly

LIST OF EXPERIMENTS

Experiment. No.	Title	Learning Outcomes
1	Study Practical: Introduction to Weka	To learn how to gather and analyze large sets of data to gain useful business understanding
2	Study Practical: Introduction to RStudio	To learn how to gather and analyze large sets of data to gain useful business understanding and how to produce a quantitative analysis report/memo with the necessary information to make Decisions
3	Data Visualization in Weka	To develop meaningful data visualizations,
4	To perform Clustering using K-Means Algorithm over Iris Dataset in Weka	To apply summary statistics techniques over datasets.
5	To Perform Frequent Pattern Mining using Apriori Algorithm over Weather.Nominal Dataset in Weka	To apply summary statistics techniques over datasets.
6	To perform Correlation Analysis over Numeric Attributes	To apply summary statistics techniques over datasets, to prepare data for analytics and perform exploratory data analysis
7	To perform Correlation Analysis over Nominal Attributes	To apply summary statistics techniques over datasets, to prepare data for analytics and perform exploratory data analysis
8	To Perform Association Rule Mining using Apriori Algorithm over Iris Dataset in RStudio	To apply summary statistics techniques over datasets.
9	To Perform Data Cleaning (Filling out missing values) using Measures of Central Tendency in RStudio	To perform cleaning and reformatting real world data for analysis, To prepare data for analytics and perform exploratory data analysis
10	To perform Linear Regression in RStudio	To apply summary statistics techniques over datasets.
11	To perform Clustering using K-Means Algorithm over Iris Dataset in RStudio	To apply summary statistics techniques over datasets.

Subject: Recommender System (Department Elective - II)								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0107			Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

Course objectives:

1. To learn techniques for making recommendations, including non-personalized, content-based, and collaborative filtering
2. To automate a variety of choice-making strategies with the goal of providing affordable, personal, and high-quality recommendations

Course Content

UNIT-I

[12 hours]

Introduction

Overview of Information Retrieval, Retrieval Models, Search and Filtering Techniques: Relevance Feedback, User Profiles, Recommender system functions, Matrix operations, covariance matrices, Understanding ratings, Applications of recommendation systems, Issues with recommender system.

UNIT-II

[12 hours]

Content-based Filtering

High level architecture of content-based systems, Advantages and drawbacks of content based filtering, Item profiles, Discovering features of documents, pre-processing and feature extraction, Obtaining item features from tags, Methods for learning user profiles, Similarity based retrieval, Classification algorithms.

Collaborative Filtering

User-based recommendation, Item-based recommendation, Model based approaches, Matrix factorization, Attacks on collaborative recommender systems.

UNIT-III

[12 hours]

Hybrid approaches

Opportunities for hybridization, Monolithic hybridization design: Feature combination, Feature augmentation, Parallelized hybridization design: Weighted, Switching, Mixed, Pipelined hybridization design: Cascade Meta-level, Limitations of hybridization strategies

Evaluating Recommender System

Introduction, General properties of evaluation research, Evaluation designs: Accuracy, Coverage, confidence, novelty, diversity, scalability, serendipity, Evaluation on historical datasets, Offline evaluations.

UNIT-IV

[12 hours]

Types of Recommender Systems

Recommender systems in personalized web search, knowledge-based recommender system, Social tagging recommender systems, Trust-centric recommendations, Group recommender systems.

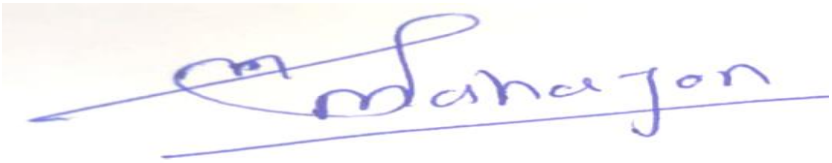
Course Outcome:

At the end of the course, students will demonstrate the ability to

1. Apply network virtualization.
2. Apply remote method invocation and objects.
3. Design process and resource management systems

Text Books:

1. Jannach D., Zanker M. and FelFering A., Recommender Systems: An Introduction, Cambridge University Press (2011), 1st ed.
2. Charu C. Aggarwal, Recommender Systems: The Textbook, Springer (2016), 1st ed.
3. Ricci F., Rokach L., Shapira D., Kantor B.P., Recommender Systems Handbook, Springer(2011), 1st ed.
4. Manouselis N., Drachsler H., Verbert K., Duval E., Recommender Systems For Learning, Springer



Subject: Machine Learning (Department Elective - II)								
Program: **M.Tech. in Data Science & Analysis**			Subject Code: DS0108				Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)-Theory**	**Continuous Internal Evaluation (CIE)-Practical**	**Total**
3	0	2	4	40	40	60	60	200

Course Outcome:

- 1. To provide deep understanding on various mathematical models and algorithms.
- 2. Learn modeling skills, optimization algorithms, and theories commonly in machine learning and data analytics.
- 3. Apply advanced statistical machine learning techniques
- 4. Understand the caveats of real-world data.
- 5. Apply modern machine learning methods in enterprise.

CONTENTS

UNIT-I

[12 hours]

Unit 1

Introduction:

Introduction and Applications of machine Learning, ML vs. DL vs. AI - What's the Difference? Types of Machine Learning Supervised, Unsupervised, Reinforcement Learning. Supervised learning can be divided further into two categories of problem: Classification, Regression, Unsupervised learning Divided in to Clustering, Association

Performance analysis of a tensor processing unit

Hardware for machine learning, The dominance of GPUs, Accelerators for machine learning Machine learning frameworks and cluster parallelism TensorFlow, SciKit, PyTorch.

UNIT-II

[12 hours]

Supervised ML Regression and Classification Techniques

Classification Algorithms

Linear Models

- Logistic Regression
- Support Vector Machines

- **Non-linear Models**
  - K-Nearest Neighbours
  - Kernel SVM
  - Naïve Bayes
  - Decision Tree Classification
  - Random Forest Classification

UNIT-III

[12 hours]

Unsupervised Learning

Clustering

- -K-means clustering,
- Hierarchical clustering

Reinforcement learning:

- Q Learning,
- Non deterministic rewards and
- Actions.

Evolutionary computing:

- Genetic Algorithms,
- Genetic Programming,
- Basic GA Operators,
- Encoding,
- Fitness Function,
- Reproduction,
- Cross Over,
- Mutation.

UNIT-IV [12 hours]

Artificial Neural Networks

Biological Neurons and Biological Neural Networks, Perceptron Learning, Activation Functions, Multilayer Perceptron, Back-propagation Neural Networks, Introduction to Deep Learning, Convolution Neural Network Relu , Pooling, Flattering, Fully Connected, Softmax

Course Outcomes

At the end of this subject, students should be able to:

- 1. Have a good understanding of the two numerical approaches to learning (optimization and integration) and how they relate to the Bayesian approach.
- 2. Have an understanding of how to choose a model to describe a particular type of data.
- 3. Know how to evaluate a learned model in practice.
- 4.Understand the mathematics necessary for constructing novel machine learning solutions.
- 5. Be able to design and implement various machine learning algorithms in a range of real-world applications.

Text Books:

1. Machine Learning Tom Mitchell; McGraw-Hill Science
2. Hands-On GPU Programming with Python and CUDA: Boost your application's performance and productivity with CUDA Brian Tuomanen

Reference Books:

1. Parallel Machine Learning Using Concurrency Control Xinghao Pan Pattern Classification Richard O. Duda, Peter E. Hart, David G. Stork
2. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods Nello Cristianini, John Shawe-Taylor; Cambridge University Press
3. Rajjan Shinghal, Pattern Recognition, Techniques and Applications, OXFORD

Web Sources: https://professional.mit.edu/programs/short-programs/advanced-machine-learning

LIST OF EXPERIMENTS

Experiment. No.	Title	Learning Outcomes
1	Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file	Understand the implementation procedures for the machine learning algorithms
2	For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples	Identify and apply Machine Learning algorithms to solve real world problems

3	Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.	Apply appropriate data sets to the Machine Learning algorithms
4	Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.	Apply appropriate data sets to the Machine Learning algorithms
5	Study parallelism in Machine Learning: GPUs, CUDA, and Practical Applications	Understand GPU-based parallelism
6	Study parallel processing in python	Learn parallelism in python
7	Write a program to demonstrate synchronous and asynchronous execution	Learn parallelism in python
8	Write a program to count how many numbers exist between a given range in each row with and without parallelization	Learn parallelism in python
9	Study how to parallelize a Pandas Data Frame	Learn parallelism in python
10	Implement multi-layer fully connected network in Keras	Develop neural network in python with keras step by step

Subject: Data Storage Technologies and Networks (Department Elective - II)								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0109			Semester**: I**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

Course objectives:

1. Evaluate storage architectures, including storage subsystems, DAS, SAN, NAS, CAS.
2. Define backup, recovery, disaster recovery, business continuity, and replication.
3. Examine emerging technologies including IP-SAN.
4. Understand logical and physical components of a storage infrastructure.
5. Identify components of managing and monitoring the data center.

Course Content

UNIT-I

[12 hours]

Storage Media and Technologies
Magnetic, Optical and Semiconductor Media, Techniques for read/write Operations, Issues and Limitations.

UNIT-II

[12 hours]

Usage and Access
Positioning in the Memory Hierarchy, Hardware and Software Design for Access, Performance issues.

UNIT-III

[12 hours]

Large Storages
Hard Disks, Networked Attached Storage, Scalability issues, networking issues.
Storage Architecture
Storage Partitioning, Storage System Design, Caching, Legacy Systems.

UNIT-IV

[12 hours]

Storage Area Networks
Hardware and Software Components, Storage Clusters/Grids. Storage QoS–Performance, Reliability, and Security issues, Recent Trends related to Copy data management, Erasure coding, and Software defined storage appliances.

Course Outcome:

At the end of the course, students will demonstrate the ability to

1. Learn Storage System Architecture
2. Ability to identify key challenges in managing information and analyze different storage networking technologies and virtualization.
3. Ability to understand components and the implementation of NAS.
4. To understand CAS architecture and types of archives and forms of virtualization.
5. To monitor the storage infrastructure and management activities.

TEXT BOOKS

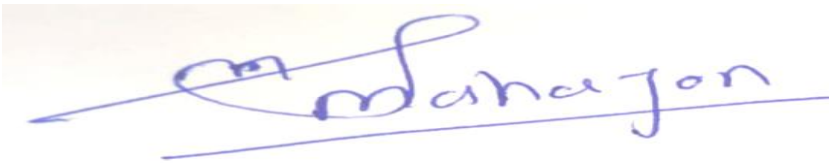
1. Storage Networking-Real World Skills for the CompTIA Storage+ Certification and Beyond by Nigel Poulton, Publishers: SYBEX a Wiley brand
2. "Storage Networks Explained" – by Ulf Troppens, Wolfgang Muller-Freidt, Rainer Wolafka, IBM Storage Software Development, Germany. Publishers: Wiley

REFERENCES

1. R The Complete Guide to Data Storage Technologies for Network-centric Computing Paperback– Import, Mar 1998 by Computer Technology Research Corporation
2. Data Storage Networking: Real World Skills for the CompTIA Storage by Nigel Poulton
3. Marc Farley, "Building Storage Networks", Tata McGraw Hill, Osborne, 2001
4. Meeta Gupta, "Storage Area Network Fundamentals", Pearson Education Limited, 2002

WEB RESOURCES

1. http://www.tsmtutorials.com/p/storage-area-network-san-basic-tutorials.html
2. https://kmitit4yr2semsanmaterial.files.wordpress.com/2017/08/unit-1.pdf
3. https://elearn.bits-pilani.ac.in



LIST OF EXPERIMENTS

Experiment. No.	Title
1	Install and configure Network Devices: HUB, Switch and Routers.
2	Connect the computers in Local Area Network.
3	Configure Host IP, Subnet Mask and Default Gateway in a System in LAN (TCP/IP Configuration).
4	Establish Peer to Peer network connection using two systems using Switch and Router in a LAN.
5	Transfer files between systems in LAN using FTP Configuration, install Print server in a LAN and share the printer in a network.
6	Study of basic network command and Network configuration commands
7	Configure a Network topology using packet tracer software
8	Installation of Proxy Server on Windows.
9	Network Troubleshooting and To create your own network cables.
10	Create a LAN in your college.
11	Implementation of file and printer sharing.
12	Designing and implementing Class A, B, C Networks Subnet planning and its implementation

SEMESTER-2

M-TECH COMPUTER SCIENCE & ENGINEERING (Data Science & Analysis) SEMESTER –II TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019													
SR NO	CODE	SUBJECTS	TEACHING SCHEME			CREDITS	HOURS	EXAMINATION SCHEME					
								THEORY		PRACT		TOTAL
			L	T	P			CIE	ESE	CIE	ESE	
1	DS0201	Advance Algorithms	3	0	2	4	5	60	40	60	40	200
2	DS0202	Soft Computing	3	0	2	4	5	60	40	60	40	200
3	DS0203	Cloud Computing	3	0	2	4	5	60	40	60	40	200
4	DS0204	Personality Enhancement	2	0	0	2	2	60	40	-	-	100
5	DS0205	Data Visualization (Elective – III)	3	0	2	4	5	60	40	60	40	200
	DS0206	Knowledge Discovery (Elective – III)										
	DS0207	Data Warehouse and Data Mining (Elective – III)										
6	DS0208	Data Security and Access Control (Elective – IV)	3	0	2	4	5	60	40	60	40	200
	DS0209	Web Analytics and Development (Elective – IV)										
	DS0210	Big Data Analytics (Elective – IV)										
TOTAL			**17**	**0**	**10**	**22**	**27**	**360**	**240**	**300**	**200**	**1100**

Subject: Advance Algorithms								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0201			Semester**: II**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

Course objectives:

1. Understand and develop a variety of techniques for designing algorithms.
2. Expand advanced algorithm analysis skills for analyzing the approximation ratio of approximation algorithms, parallel algorithm and the probability of randomized algorithms.
3. Gain a high-quality understanding on a wide range of advanced algorithmic problems, their relations and variants, and application to real-world problems.

Course Content

UNIT-I

[12 hours]

Introduction:
Basic concept of algorithm, Overview of Divide and Conquer, Greedy Algorithms and Dynamic Programming. Basic search and traversal techniques for graphs, Backtracking, Branch and Bound.

UNIT-II

[12 hours]

Maximum Flow:
Flow Networks, Ford-Fulkerson method.
String Matching
Introduction to string-matching problem, Naïve string matching algorithm, Rabin Karp, Knuth Morris Pratt, Boyer-Moore matching algorithms and complexity analysis.
NP- Hard and NP-Complete Problems:
P, NP and NP-Complete complexity classes, Proof of NP-Completeness.

UNIT-III

[12 hours]

Approximation Algorithms:
Introduction, Combinatorial Optimization, approximation factor, Types of approximation algorithm, different examples of approximation algorithm.
.**Parallel Algorithms:**

Introduction, Classification of Parallel System, PRAM Model, parallel algorithm specifications and analysis, Parallel Searching and Parallel Sorting.

UNIT-IV

[12 hours]

Probabilistic Algorithms & Randomized Algorithms
Numerical probabilistic algorithms, Las Vegas and Monte Carlo algorithms, Game-theoretic techniques.

Course outcomes:

At the end of this subject, students should be able to:

1. Analyze a variety of algorithms with practical applications and the resource requirements of every one.
2. Establish the most suitable algorithm for any given task and then apply it to the problem.
3. Demonstrate sufficient comprehension of the theory of intractability and prove when certain kinds of problems are intractable.

TEXT BOOKS

1. Norton M Introduction to Algorithms: T.H. Cormen, C.E.Leiserson and R.L. Rivest
2. Randomized Algorithms: R. Motwani and P.Raghavan

REFERENCES

1. Fundamentals of Algorithmic : G. Brassard and P. Bratley
2. Approximation Algorithms: Vijay V. Vazirani
3. Reference book: Algorithmic :The spirit of computing: D. Harel
4. Design and analysis of algorithms, S. Sridhar

WEB RESOURCES

https://www.geeksforgeeks.org/randomized-algorithms/
https://www.tutorialspoint.com/parallel_algorithm/parallel_algorithm_introduction.htm

LIST OF EXPERIMENTS

Experiment. No.	Title
1.1	Implement algorithm and program for merge sort using divide and conquer strategy.
1.2	Implement algorithm and program for quick sort using divide and conquer strategy.
1.3	Implement program for minimum spanning tree using greedy technique.
1.4	Implement program for knapsack problem using dynamic programming.
2.1	Implementation of Ford Fulkerson algorithm.
2.2	Use following algorithm for string matching: 1. String Matching 2. Naïve String Matching 3. Rabin Karp
2.3.	Implement program for Knuth Morris Pratt technique.
3.1	Implement program for Enumeration Sort using parallel algorithm
3.2	Implement program for Odd-Even Transposition Sort using parallel algorithm
3.3	Perform Travelling Salesman problem using approximation algorithm.
4.1	Implement program for randomized quick sort.
4.2	Case Study

Subject: Soft Computing								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0202			Semester**: II**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

Course objectives:

1. Soft computing refers to principle components like fuzzy logic, neural networks and genetic algorithm, which have their roots in Artificial Intelligence.
2. Healthy integration of all these techniques has resulted in extending the capabilities of the technologies to more effective and efficient problem solving methodologies

Course Content

UNIT-I

[12 hours]

Introduction to Neural Network

Biological Neuron, Fundamental concept of Artificial Neural Network & Biological Neural Network , Evolution of Neural Networks, Basic Models of Artificial Neural Networks ,Types of Learning, Important Terminologies of ANNs ,Idea of computational units & Activation Functions, McCulloch–Pitts unit and Thresholding logic, Linear separability, Hebb network

Supervised Learning Networks

Perceptron, Perceptron Learning Algorithm, Convergence theorem for Perceptron Learning Algorithm, Adaline ,Multiple Adaptive Linear Neurons , Multi-layered Networks, Generalized delta rule, Gradient Descent, Backpropagation Network, Radial basis function network, Adaptive Learning rate, Feedback Networks, Back-propagation through time, real-time recurrent learning.

UNIT-II

[12 hours]

Unsupervised Learning Networks

Competitive Learning, Fixed weight Competitive Nets, Kohonen Self-organizing Feature Maps, Extended Kohonen SOM.

Associative Memory Networks

Training Algorithms for Pattern Association, Auto associative Memory Network , Heteroassociative Memory Network , Bidirectional Associative Memory , Hopfield Networks , Iterative Autoassociative Memory Networks , Temporal Associative Memory Network

UNIT-III

[12 hours]

Fuzzy Logic

Basic concepts of fuzzy logic, Fuzzy sets and Crisp sets, Fuzzy set theory and operations, Properties of fuzzy sets, Fuzzy and Crisp relations, Fuzzy to Crisp conversion. Fuzzy Logic –II (Fuzzy Membership, Rules) Membership functions, interference in fuzzy logic, fuzzy if-then rules, Fuzzy implications and Fuzzy algorithms, Fuzzifications & Defuzzification, Fuzzy Controller, Industrial applications.

UNIT-IV

[12 hours]

Genetic Algorithm(GA)

Basic concepts, working principle, procedures of GA, flow chart of GA, Genetic representations, (encoding) Initialization and selection, Selection techniques, Crossover Techniques, ,Genetic operators, Mutation, Mutation techniques ,Problem Solving using GA, applications.

Course Outcomes:

The students should be able to:

1. Identify and describe soft computing techniques and their roles in building intelligent machines.
2. Recognize the feasibility of applying a soft computing methodology for a particular problem.
3. Apply fuzzy logic and reasoning to handle uncertainty and solve engineering problems.
4. Apply genetic algorithms to combinatorial optimization problems.
5. Apply neural networks to pattern classification and regression problems.
6. Effectively use existing software tools to solve real problems using a soft computing approach.

TEXT BOOKS

1. S.N. Sivanandan and S.N. Deepa, Principles of Soft Computing, Wiley India, 2007. ISBN: 10: 81-265-1075-7.
2. S. Rajsekaran & G.A. Vijayalakshmi Pai, “Neural Networks,Fuzzy Logic and Genetic Algorithm:Synthesis and Applications” Prentice Hall of India.

REFERENCES

1. Siman Haykin,”Neural Netowrks”Prentice Hall of India
2. Timothy J. Ross, “Fuzzy Logic with Engineering Applications” Wiley India.
3. Kumar Satish, “Neural Networks” Tata Mc Graw Hill

REFERENCES

https://www.youtube.com/watch?v=aYf3MKzExzI
https://www.youtube.com/watch?v=0kgEjAAjxM8
https://www.youtube.com/watch?v=HfdZE1dogfA
https://www.youtube.com/watch?v=Z_8MpZeMdD4

LIST OF EXPERIMENTS

Exp. No.	Title	Learning Outcomes
1	W.A.P to implement simple AND & OR function with neuron.(take w1=w2=1, find appropriate threshold(threshold)	Basic knowledge of neuron model.
2	Write a program to generate XOR function using McCulloch-Pitts neuron and appropriate values for weights, bias and threshold.	Basic knowledge of neuron model.
3	Write a program for perceptron net for an AND function with bipolar inputs and targets.	Basic knowledge of perceptron learning algorithm.
4	Write a program to Implement Hebb Algorithm.	Basic knowledge of hebb algorithm in In neural network.
5	Write a program to Implement ADALINE Algorithm.	Basic knowledge of ADALINE algorithm in In neural network.
6	Write a program to Implement MADALINE Algorithm.	Basic knowledge of MADALINE algorithm in In neural network.
7	Write a program to Implement BPN(Back Propagation Network) Algorithm.	Basic knowledge of back propogation algorithm in In neural network.
8	W.A.P to implement fuzzy set properties & operations.	Basic knowledge of fuzzy logic.
9	Solve a given problem-1(Operatons) using Fuzzy Logic in MATLAB.	Basic knowledge of fuzzy logic.
10	Solve a given problem (Max-Min Composition) using Fuzzy Logic in MATLAB.	Basic knowledge of fuzzy logic.

Subject: Cloud Computing								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0203			Semester**: II**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
2	0	2	3	40	40	60	60	200

Course objectives:

1. Identify the technical foundations of Cloud systems architecture.
2. Analyze the problems and solutions to cloud application problems.
3. Identify the research scope in cloud computing.
4. Describe various service delivery models of cloud computing architecture, and the ways in which clouds can be deployed as public, private, hybrid, and community clouds.
5. Comprehend the technical capabilities and business benefits of virtualization and cloud computing.
6. Describe the landscape of different types of virtualization and understand the different types of clouds.

Course Content

UNIT-I

[12 hours]

Introduction to Cloud Computing

Cloud Computing Overview, History of Cloud Computing, Layers and Types of Cloud, Offerings of a cloud, Software-as-a-Service, Platform-as-a-Service, Infrastructure-as-a-Service, Challenges and Risks.

Cloud Computing Architecture and Vitalization

Cloud Computing Architecture, Deployment Models, Virtualization, XML Basics, web Services, Service Oriented Architecture.

UNIT-II

[12 hours]

Managing a Cloud and SLA

Managing cloud data, Introduction to MapReduce, OpenStack, Cloud Economics, Service Level Agreement (SLA), Resource Management, Case Studies Commercial Cloud and Google Cloud Platform.

Virtualization of the resource provisioning

Virtual machine technology, virtualization applications in enterprises, Drawbacks of virtualization.

Multitenancy on offering

Multi-entity support, Multi-schema approach, Multitenancy using cloud data stores, Data access control for enterprise applications.

UNIT-III

[12 hours]

Cloud Security Aspects

Cloud security fundamentals

Vulnerability assessment tool for cloud, Privacy and Security in cloud

Cloud computing security Structure

Architectural Considerations- General Issues, Trusted Cloud computing, Secure Execution Environments and Communications, Micro-architectures; Identity Management and Access control Identity management, Access control, Autonomic Security, Virtualization security management virtual threats, VM Security Recommendations, VM-Specific Security techniques.

Cloud computing security Issues

Cloud Computing: Security Issues in Collaborative SaaS Clo, Cloud Computing: Broker for Cloud Marketplace

UNIT-IV

[12 hours]

Recent trends and Research scope in cloud computing

Mobile cloud Computing, Fog Computing, Geo-Spatial cloud, Green Cloud Computing, IoT Cloud, Big Data and Cloud Computing, Introduction to Docket Container, Research Scope of the Cloud Computing, Open Source and Commercial Clouds, Cloud Simulator.

Course Outcomes:

The students should be able to:

1. Explore the research scope of cloud computing.
2. Implement the virtual cloud.
3. Security perusal of data in cloud environment.
4. Compare various cloud service provider architecture.

TEXT BOOKS

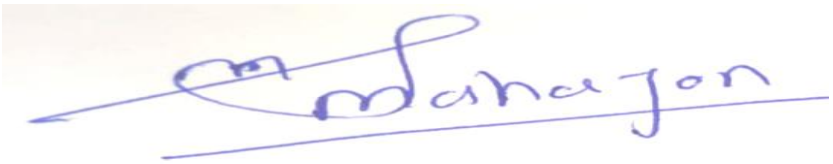
1. Rajkumar Buyya, Cloud Computing: Principles and Paradigms, John Wiley & Sons, First Edition
2. Gautam Shroff, Enterprise Cloud Computing: Technology, Architecture, Applications, Cambridge University Press, First EditionJudith Hurwitz, R Bloor, M.Kanfman, F.Halper, Cloud Computing for Dummies, Wiley India Edition, First Edition

REFERENCES

1. Anthony T Velte, Cloud Computing : A Practical Approach, McGraw-Hill Osborne, First Edition
2. Barrie Sosinsky, Cloud Computing Bible, Wiley India, First Edition
3. Ronald Krutz and Russell Dean Vines, Cloud Security, Wiley-India, First Edition
4. Tim Malhar, S.Kumaraswammy, S.Latif, Cloud Security & Privacy, O'Really Publications, First Edition
5. Scott Granneman, Google Apps, Pearson, First Edition

WEB RESOURCES

http://cloudbus.org
http://www.salesforce.com/in/cloudcomputing/
http://www.ibm.com/cloud-computing/in/en/what-is-cloud-computing.html
http://www.rackspace.com/cloud/what_is_cloud_computing
http://aws.amazon.com/
http://www.microsoft.com/en-in/server-cloud/cloud-os/
http://azure.microsoft.com/en-in/
https://cloud.google.com/
https://cloud.google.com/products/



LIST OF EXPERIMENTS

Ex. No.	Title	Learning Outcomes
1.1	Sketch out and analyze architecture of Cloudsim and identify different entities to understand the structure of cloudsim.	Usage of CLoudSim and Architecture of the same.
2.1	Create a scenario in cloudsim to create a datacenter along with one host. Also create one virtual machine with static configuration to run one cloudlet on it.	Working of the data centers in cloud environment.
2.2	Illustrate a scenario in cloudsim to create one datacenter and one host. Also implement required virtual machines to run two cloudlets on it. Assume that cloudlets run in VMs with the same MIPS requirements. The cloudlets will take the same time to complete the execution.	Understanding the cloudlets request in a cloud scenario.
3.1	Implement a datacenter with two hosts and run two cloudlets on it in cloudsim. Consider the cloudlets run in VMs with different MIPS requirements. The cloudlets will take different time to complete the execution depending on the requested VM performance.	Analyzing the cloud performance in the distinct scenario.
3.2	Design a program in cloudsim to create two datacenters with one host and run two cloudlets on it.	Multiple cloudlet execution in the cloud.
4.1	Construct a case in cloudsim to create two datacenters with one host each and run cloudlets of two users on them.	Multiple data center approach and its performance on the cloud.
5.1	Make and perform scenario to pause and resume the simulation in cloudsim, and create simulation entities (a DatacenterBroker) dynamically	Understanding the Brocker concept with example.
5.2	Organize a case in cloudsim for simulation entities (a DatacenterBroker) in run-time using a globar manager entity (GlobalBroker).	Understanding the global manager in cloud.
6.1	Sketch out and analyze architecture of Microsoft Azure.	Working of AZURE cloud.
7.1	Implement a web application using Microsoft Azure account as a cloud service by creating a web page and database. Also provide database connectivity with implemented webpage.	Utilizing the cloud services.

Subject: **Personality Enhancement**

Program: **M.Tech Personality Enhancement**			Subject Code:DS0204		Semester: II			
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
2	0	0	2	40	0	60	0	100

Course Objectives:

1. To enhance holistic development of students and improve their employability skills
2. To acquire inter personal skills and be an effective goal oriented team player
3. To develop professionalism with idealistic, practical and moral values
4. To acquire communication and problem solving skills
5. To re-engineer their attitude and understand and understand its influence on behavior

COURSE CONTENT

UNIT I

ATTITUDE: Who am I? SWOT analysis, Importance of self-confidence and self-esteem, Factors influencing attitude

UNIT II

COMMUNICATION: Verbal, Non-verbal, Vocal, Activities for evaluation (Extempore, speaking news, book review), Public Speaking Activities for evaluation (Surveying and reporting, Debate, Group discussion)

UNIT III

GOAL SETTING: SMART goals, Blue print for success, Short term, Long term, Life time goals, Value of time, Diagnosing time management, Prioritizing work, Factors of motivation, Self-talk, Intrinsic & Extrinsic Motivators

UNIT IV

ENGINEERING ETHICS: Ethical Reasoning, Responsibilities towards environment and society, Leading by example, Concern for quality, Concern for success

Recommended Reading

- Covey Sean, Seven Habits of Highly Effective Teens, New York, Fireside Publishers, 1998.
- Carnegie Dale, How to win friends and influence people, New York, Simon and Schuster, 1998
- Thomas A Harris, I am ok, you are ok, New York, Harper and Row, 1972
  Daniel Coleman, Emotional Intelligence, Bantam Book, 2006Ethics in Engineering Practice and Research by Caroline &Whitbeck, Cambridge University Press

Subject: Data Visualization (Elective – III)		
Program: **M.Tech. in Data Science & Analysis**	Subject Code: DS0205	Semester**: II**

Teaching Scheme				Examination Evaluation Scheme				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

COURSE OBJECTIVE

1. familiarize students with the basic and advanced techniques of information visualization and scientific visualization,
2. to learn key techniques of the visualization process a detailed view of visual perception, the visualized data and the actual visualization, interaction and distorting techniques.

Course Content

Unit I

[16 hours]

Introduction of visual perception, visual representation of data, Gestalt principles, information overloads.

Creating visual representations, visualization reference model, visual mapping, visual analytics, Design of visualization applications.

Unit II

[10 hours]

Classification of visualization systems, Interaction and visualization techniques misleading, Visualization of one, two and multi-dimensional data, text and text documents.

Unit III

[11 hours]

Visualization of groups, trees, graphs, clusters, networks, software, Metaphorical visualization

Unit IV

[11 hours]

Visualization of volumetric data, vector fields, processes and simulations, Visualization of

maps, geographic information, GIS systems, collaborative visualizations, Evaluating visualizations

Recent trends in various perception techniques, various visualization techniques, data structures used in data visualization.

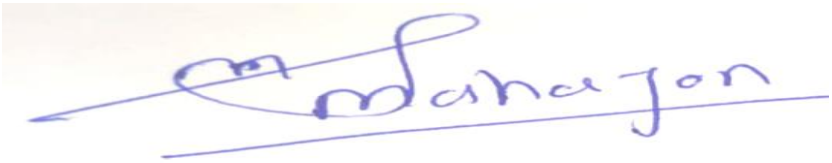
COURSE OUTCOMES

On completion of the course the student should be able to

1. Familiar with the design process to develop visualization methods and visualization systems, and methods for their evaluation.
2. preparation and processing of data, visual mapping and the visualization
3. have an understanding of large-scale abstract data,

References:

1. WARD, GRINSTEIN, KEIM,.Interactive Data Visualization: Foundations, Techniques, and Applications. Natick : A K Peters, Ltd.
2. E. Tufte, The Visual Display of Quantitative Information, Graphics Press



Subject: Knowledge Discovery (Elective – III)		
Program: **M.Tech. in Data Science & Analysis**	Subject Code: DS0206	Semester**: II**
Teaching Scheme	**Examination Evaluation Scheme**	

Lecture	Tutorial	Practical	Credits	University Theory Examination	University Practical Examination	Continuous Internal Evaluation (CIE)- Theory	Continuous Internal Evaluation (CIE)- Practical	Total
3	0	2	4	40	40	60	60	200

COURSE OBJECTIVE

1. Conduct case studies on real data mining examples

Unit I

[17 HOUR]

Introduction KDD and Data Mining - Data Mining and Machine Learning, Machine Learning and Statistics, Generalization as Search,Data Mining and Ethics

Knowledge Representation - Decision Tables, Decision Trees, Classification Rules, Association Rules, Rules involving Relations, Trees for Numeric Predictions, Neural Networks, Clusters

Unit II

[10 HOUR]

Decision Trees - Divide and Conquer, Calculating Information, Entropy, Pruning, Estimating Error Rates, The C4.5 Algorithm

Evaluation of Learned Results- Training and Testing, Predicting Performance, Cross-Validation

Unit III

[15 HOUR]

Unit 4:

Classification Rules - Inferring Rudimentary Rules, Covering Algorithms for Rule Construction, Probability Measure for Rule Evaluation, Association Rules, Item Sets, Rule Efficiency

Numeric Predictions - Linear Models for Classification and Numeric Predictions, Numeric Predictions with Regression Trees, Evaluating Numeric Predictions

Unit IV

[14 HOUR]

Artificial Neural Networks – Perceptrons, Multilayer Networks, The Backpropagation Algorithm

Clustering - Iterative Distance-based Clustering, Incremental Clustering, The EM Algorithm

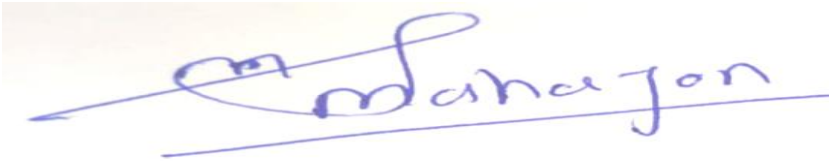
COURSE OUTCOMES

After completion of course, students would be:

Able to have knowledge of various knowledge representation methods.

References:

1. Data mining and knowledge discovery handbook by Maimon, oded(et al.)
2. Data Cleansing : A Prelude to knowledge Discovery



Subject: Data Warehouse and Data Mining (Elective – III)			
Program: **M.Tech. in Data Science & Analysis**	Subject Code:	Semester**: II**	
Teaching Scheme	**Examination Evaluation Scheme**		

Lecture	Tutorial	Practical	Credits	University Theory Examination	University Practical Examination	Continuous Internal Evaluation (CIE)- Theory	Continuous Internal Evaluation (CIE)- Practical	Total
3	0	2	4	30/60	30/60	20/40	20/40	200

COURSE OBJECTIVE

1. The objective of this course is to introduce data warehousing and mining techniques.
2. Application of data mining in web mining, pattern matching and cluster analysis is included
3. to aware students of broad data mining areas.

Course Content

UNIT-I

[15 hours]

Introduction to Data Warehousing; Data Mining: Mining frequent patterns, association and correlations; Sequential Pattern Mining concepts, primitives, scalable methods;
Classification and prediction; Cluster Analysis – Types of Data in Cluster Analysis, Partitioning methods, Hierarchical Methods; Transactional Patterns and other temporal based frequent patterns,

UNIT-II

[18 hours]

Mining Time series Data, Periodicity Analysis for time related sequence data, Trend analysis, Similarity search in Time-series analysis;
Mining Data Streams, Methodologies for stream data processing and stream data systems, Frequent pattern mining in stream data, Sequential Pattern Mining in Data Streams, Classification of dynamic data streams, Class Imbalance Problem; Graph Mining; Social Network Analysis;
11

UNIT-IV

[18 hours]

Web Mining, Mining the web page layout structure, mining web link structure, mining

multimedia data on the web, Automatic classification of web documents and web usage mining; Distributed Data Mining.

Recent trends in Distributed Warehousing and Data Mining, Class Imbalance Problem; Graph Mining; Social Network Analysis

COURSE OUTCOMES

After completion of course, students would be:

1. Study of different sequential pattern algorithms
2. Study the technique to extract patterns from time series data and it application in real world.
3. Can extend the Graph mining algorithms to Web mining
4. Help in identifying the computing framework for Big Data

References:

1. Jiawei Han and M Kamber, Data Mining Concepts and Techniques,, Second Edition, Elsevier Publication, 2011.
2. Vipin Kumar, Introduction to Data Mining - Pang-Ning Tan, Michael Steinbach, Addison Wesley,2006.
3. G Dong and J Pei, Sequence Data Mining, Springer, 2007.

Subject: Data Security and Access Control (Elective – IV)								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0208			Semester**: II**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	4	40	40	60	60	200

COURSE OBJECTIVE

- The objective of the course is to provide fundamentals of database security. Various access control techniques mechanisms were introduced along with application areas of access control techniques.

Course Content

UNIT-I

[9 hours]

Introduction to Access Control, Purpose and fundamentals of access control, brief history, Policies of Access Control, Models of Access Control, and Mechanisms, Discretionary Access Control (DAC), Non- Discretionary Access Control, Mandatory Access Control (MAC). Capabilities and Limitations of Access Control Mechanisms: Access Control List (ACL) and Limitations, Capability List and Limitations.

UNIT-II

[18 hours]

Role-Based Access Control (RBAC) and Limitations, Core RBAC, Hierarchical RBAC, Statically Constrained RBAC, Dynamically Constrained RBAC, Limitations of RBAC.
Comparing RBAC to DAC and MAC Access control policy.
Biba'sintrigity model, Clark-Wilson model, Domain type enforcement model, mapping the enterprise view to the system view, Role hierarchies- inheritance schemes, hierarchy structures and inheritance forms, using SoD in real system Temporal Constraints in RBAC, MAC AND DAC. Integrating RBAC with enterprise IT infrastructures: RBAC for WFMSs, RBAC for UNIX and JAVA environments Case study: Multi-line Insurance Company

UNIT-III

[11 hours]

Smart Card based Information Security, Smart card operating system fundamentals, design and implantation principles, memory organization, smart card files, file management, atomic operation, smart card data transmission ATR, PPS Security techniques- user

identification, smart card security, quality assurance and testing, smart card life cycle-5 phases, smart card terminals.

UNIT-IV

[11 hours]

Recent trends in Database security and access control mechanisms. Case study of Role-Based Access Control (RBAC) systems.
Recent Trends related to data security management, vulnerabilities in different DBMS.

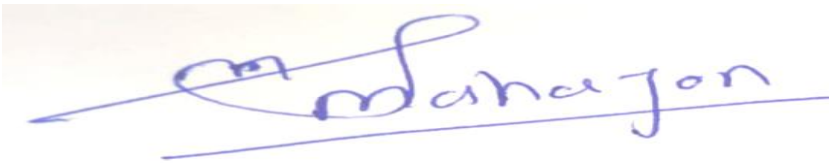
COURSE OUTCOMES

After completion of course, students would be:

1. In this course, the students will be enabled to understand and implement classical models and algorithms
2. They will learn how to analyze the data, identify the problems, and choose the relevant models and algorithms to apply.
3. They will further be able to assess the strengths and weaknesses of various access control models and to analyze their behavior.

References:

1. Role Based Access Control: David F. Ferraiolo, D. Richard Kuhn, RamaswamyChandramouli.
2. http://www.smartcard.co.uk/tutorials/sct-itsc.pdf: Smart Card Tutorial.



Subject: Big Data Analytics (Department Elective - IV)								
Program: **M.Tech. in Data Science & Analysis**				Subject Code: DS0210			Semester**: II**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**

3	0	2	4	40	40	60	60	200

Course objectives:

1. To understand big data for business intelligence.
2. To learn business case studies for big data analytics.
3. To understand nosql big data management. Perform map-reduce analytics using Hadoop and related tools

Course Content

UNIT-I

[8 hours]

What is big data, why big data, convergence of key trends, unstructured data, industry examples of big data, web analytics, big data and marketing, fraud and big data, risk and big data, credit risk management, big data and algorithmic trading, big data and healthcare, big data in medicine, advertising and big data, big data technologies, introduction to Hadoop, open source technologies, cloud and big data, mobile business intelligence, Crowd sourcing analytics, inter and trans firewall analytics.

UNIT-II

[8 hours]

Introduction to NoSQL, aggregate data models, aggregates, key-value and document data models, relationships, graph databases, schemaless databases, materialized views, distribution models, sharding, master-slave replication, peer-peer replication, sharding and replication, consistency, relaxing consistency, version stamps, map-reduce, partitioning and combining, composing map-reduce calculations.

UNIT-III

[19 hours]

Data format, analyzing data with Hadoop, scaling out, Hadoop streaming, Hadoop pipes, design of Hadoop distributed file system (HDFS), HDFS concepts, Java interface, data flow, Hadoop I/O, data integrity, compression, serialization, Avro, file-based data structuresMapReduce workflows, unit tests with MRUnit, test data and local tests, anatomy of MapReduce job run, classic Map-reduce, YARN, failures in classic Map-reduce and YARN, job scheduling, shuffle and sort, task execution, MapReduce types, input formats, output formats

UNIT-IV

[9 hours]

Hbase, data model and implementations, Hbase clients, Hbase examples, praxis.Cassandra, Cassandra data model, Cassandra examples, Cassandra clients, Hadoop integration.Pig, Grunt, pig data model, Pig Latin, developing and testing Pig Latin scripts. Hive, data types and file formats, HiveQL data definition, HiveQL data manipulation, HiveQL queries.

Course Outcomes:

The students should be able to:

1. Describe big data and use cases from selected business domains
2. Explain NoSQL big data management
3. Install, configure, and run Hadoop and HDFS
4. Perform map-reduce analytics using Hadoop
5. Use Hadoop related tools such as HBase, Cassandra, Pig, and Hive for big data analytics

TEXT BOOKS

1. Michael Minelli, Michelle Chambers, and AmbigaDhiraj, "Big Data, Big Analytics: Emerging
2. Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.
3. P. J. Sadalage and M. Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of

REFERENCES

1. Patankar, S.V., "Numeric Polyglot Persistence", Addison-Wesley Professional, 2012.
2. Tom White, "Hadoop: The Definitive Guide", Third Edition, O'Reilley, 2012.
3. Eric Sammer, "Hadoop Operations", O'Reilley, 2012.
4. E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilley, 2012.
5. Lars George, "HBase: The Definitive Guide", O'Reilley, 2011.
6. Eben Hewitt, "Cassandra: The Definitive Guide", O'Reilley, 2010.
7. Alan Gates, "Programming Pig", O'Reilley, 2011.

WEB RESOURCES

https://onlinecourses.nptel.ac.in/noc19_cs33/preview
https://www.coursera.org/courses?query=big%20data%20analytics

LIST OF EXPERIMENTS

ExNo.	Title	Learning Outcomes
1	Study Practical: Introduction to Hadoop and its Functionalities.	Students will be able to understand the Hadoop and its Functionalities
2	Perform Basic Hadoop shell commands to manage HDFS. (with theory)	Students will be able to understand the Hadoop Shell Commands.
3	Study Practical: Intro to NoSQL and MongoDB	Students will be able to understand NoSQL and MongoDB
4	Study Practical: MapReduce	Students will be able to understand MapReduce
5	Write a program to calculate frequency of each word using MapReduce.	Students will be able to understand how to to calculate frequency of each word using MapReduce.
6	Write a program to know the lifetime value of each customer using MapReduce Reduce Side Join.	Students will be able to understand MapReduce Reduce Side Join
7	Write a program to perform basic HBase commands.	Students will be able to understand basic HBase commands
8	Write a program to data of file in HBase table.	Students will be able to understand HBase table.
9	Write a program using Pig find the most occurred start letter.	Students will be able to understand pig.
10	Write a program in pig script to read data from a data file and to display the required contents on the terminal as output.	Students will be able to understand pig script.

Subject: Web Analytics and Development (Department Elective-IV)								
Program: **M.Tech. Automotive Engineering**				Subject Code: DS0209			Semester**: II**	
Teaching Scheme				**Examination Evaluation Scheme**				
Lecture	**Tutorial**	**Practical**	**Credits**	**University Theory Examination**	**University Practical Examination**	**Continuous Internal Evaluation (CIE)- Theory**	**Continuous Internal Evaluation (CIE)- Practical**	**Total**
3	0	2	3	40	40	60	60	200

Course objectives:

1. The course explores use of social network analysis to understand growing connectivity and complexity in the world ranging from small groups to WWW.
2. Learn the information retrieval models.
3. Be familiar with Web Search Engine.
4. Be exposed to Link Analysis.
5. Understand Hadoop and Map Reduce.
6. Learn document text mining techniques

Course Content

UNIT-I

[12 hours]

Introduction

Social network and Web data and methods, Graph and Matrices, Basic measures for individuals and networks, Information Visualization

UNIT-II

[12 hours]

Web Analytics tools

Click Stream Analysis, A/B testing, Online Surveys

Web Search and Retrieval

Search Engine Optimization, Web Crawling and indexing, Ranking Algorithms, Web traffic models

UNIT-III

[12 hours]

Making Connection

Link Analysis, Random Graphs and Network evolution, Social Connects: Affiliation and identity

UNIT-IV

[12 hours]

Connection

Connection Search, Collapse, Robustness Social involvements and diffusion of innovation

Course Outcomes:

Upon completion of this course the student will be able to:

1. Become familiar with core research communities, publications, focused on web and social media analytics and research questions engaged in
2. Design Web Search Engine.
3. Use Link Analysis.
4. Use Hadoop and Map Reduce.
5. Apply document text mining technique

TEXT BOOKS

1. Hansen, Derek, Ben Sheiderman, Marc Smith. 2011. Analyzing Social Media Networks with NodeXL: Insights from a Connected World. Morgan Kaufmann, 304.

2. Avinash Kaushik. 2009. Web Analytics 2.0: The Art of Online Accountability.

REFERENCES

1. Easley, D. & Kleinberg, J. (2010). Networks, Crowds, and Markets: Reasoning About a Highly Connected World. New York: Cambridge University Press. http://www.cs.cornell.edu/home/kleinber/networks-book/

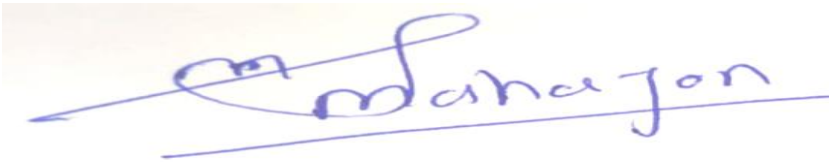
2. Wasserman, S. & Faust, K. (1994). Social network analysis: Methods and applications. New York: Cambridge University Press. Monge, P. R. & Contractor, N. S. (2003). Theories of communication networks. New York: Oxford University Press.

WEB RESOURCES

1.https://www.javatpoint.com/how-search-engine-works
2.https://www.geeksforgeeks.org/google-search-works/
3.https://nlp.stanford.edu/IR-book/pdf/20crawl.pdf
4.https://moz.com/beginners-guide-to-seo/how-search-engines-operate

LIST OF EXPERIMENTS

Experiment. No.	Title
1	Implement Search engine optimization on website using google analytics
2	Implement Search engine optimization on personal blogs using google analytics
3	Implement Search engine optimization on website using SEMRUSH.
4	Implement Search engine optimization on personal blogs using SEMRUSH.
5	Implement Search engine optimization on website using small SEO tools
6	Implement Search engine optimization on personal blogs using small SEO tools
7	Implement Search engine optimization on website using SEO analyzer
8	Implement Search engine optimization on personal blogs using SEO analyzer.
9	Study about documents mining techniques.
10	Implements map reducing using hadoop



M. TECH COMPUTER ENGINEERING (DATA SCIENCE & ANALYSIS) SEMESTER –III TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019														
SR NO	CODE	SUBJECTS	TEACHING SCHEME			CREDITS	HOURS	EXAMINATION SCHEME						
								THEORY			PRACT		TOTAL
								CIE		ESE	CIE	ESE	
			L	T	P			MID	IE				
2		Dissertation Phase-I	0	0	24	12	24	00	00	00	60	40	100
TOTAL			0	0	24	12	24	00	00	00	60	40	100

M. TECH COMPUTER ENGINEERING (DATA SCIENCE & ANALYSIS)SEMESTER –III TEACHING & EXAMINATION SCHEME WITH EFFECT FROM JULY 2019														
SR NO	CODE	SUBJECTS	TEACHING SCHEME			CREDITS	HOURS	EXAMINATION SCHEME						
								THEORY			PRACT		TOTAL
								CIE		ESE	CIE	ESE	
			L	T	P			MID	IE				
1		Dissertation Phase-II	0	0	24	12	24	00	00	00	60	40	100
TOTAL			**0**	**0**	24	12	24	00	00	00	60	40	100