ANNA UNIVERSITY, CHENNAI
NON - AUTONOMOUS COLLEGES AFFILIATED ANNA UNIVERSITY
M.E. BIG DATA ANALYTICS

REGULATIONS – 2021
CHOICE BASED CREDIT SYSTEM

1. **PROGRAMME EDUCATIONAL OBJECTIVES (PEOs):**

I. Apply quantitative modeling and data analysis techniques to the solution of real world business problems, communicate findings, and effectively present results using data visualization techniques
II. Apply principles of Data Science to the analysis of business problems.
III. Employ cutting edge tools and technologies to analyze big data.
IV. Adopt ethical practices, collaborate with team members as a member or leader, and engage in constant updation of tools and techniques.
V. Strongly focus on ingenious ideas and critical analysis to serve the society, locally and internationally as entrepreneurs using the principles of data analytics.

2. PROGRAM OUTCOMES (POs):

1. An ability to independently carry out research / investigation and development work to solve practical problems.
2. An ability to write and present a substantial technical report/document.
3. Students should be able to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program.
4. Understand the impact of big data for business decisions and strategy.
5. Gain hands-on experience on large-scale analytics tools to solve some open big data problems.
6. Understand the concept and challenge of big data and the inadequacy of existing technologies to analyze big data.

ANNA UNIVERSITY, CHENNAI
NON - AUTONOMOUS COLLEGES AFFILIATED ANNA UNIVERSITY
M.E. BIG DATA ANALYTICS
REGULATIONS – 2021
CHOICE BASED CREDIT SYSTEM
I TO IV SEMESTERS CURRICULA AND SYLLABI
SEMESTER I

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
THEORY								
1.	MA4109	Applied Probability and Statistics	FC	4	0	0	4	4
2.	RM4151	Research Methodology and IPR	RMC	2	0	0	2	2
3.	CP4151	Advanced Data Structures and Algorithms	PCC	3	0	0	3	3
4.	BD4151	Foundations of Data Science	PCC	3	0	0	3	3
5.	CP4252	Machine Learning	PCC	3	0	2	5	4
6.	CP4152	Database Practices	PCC	3	0	2	5	4
7.		Audit Course – I*	AC	2	0	0	2	0
PRACTICALS								
8.	CP4161	Advanced Data Structures and Algorithms Laboratory	PCC	0	0	4	4	2
9.	BD4111	Big Data Computing Laboratory	PCC	0	0	2	2	1
			TOTAL	**20**	**0**	**10**	**30**	**23**

*Audit course is optional

SEMESTER II

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
THEORY								
1.	BD4201	Big Data Security	PCC	3	0	0	3	3
2.	BD4251	Big Data Mining and Analytics	PCC	3	0	0	3	3
3.	MP4251	Cloud Computing Technologies	PCC	3	0	0	3	3
4.	BD4202	Information Storage Management	PCC	3	0	0	3	3
5.	MP4151	Embedded Systems and IIOT	PCC	3	0	2	5	4
6.		Professional Elective I	PEC	3	0	0	3	3
7.		Audit Course – II*	ACC	2	0	0	2	0
PRACTICALS								
8.	BD4211	Big Data Mining And Analytics Laboratory	PCC	0	0	4	4	2
9.	BD4212	Term Paper Writing and Seminar	EEC	0	0	2	2	1
			TOTAL	**20**	**0**	**8**	**28**	**22**

*Audit course is optional

SEMESTER III

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
THEORY								
1.		Professional Elective II	PEC	3	0	0	3	3
2.		Professional Elective III	PEC	3	0	0	3	3
3.		Professional Elective IV	PEC	3	0	0	3	3
4.		Open Elective	OEC	3	0	0	3	3
PRACTICALS								
5.	BD4311	Project Work I	EEC	0	0	12	12	6
			TOTAL	**12**	**0**	**12**	**24**	**18**

SEMESTER IV

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
PRACTICALS								
1.	BD4411	Project Work II	EEC	0	0	24	24	12
			TOTAL	**0**	**0**	**24**	**24**	**12**

TOTAL NO. OF CREDITS: 75

PROFESSIONAL ELECTIVES
SEMESTER II, ELECTIVE I

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	BD4071	High Performance Computing for Big Data	PEC	3	0	0	3	3
2.	MP4094	Web Services and API Design	PEC	3	0	0	3	3
3.	CP4093	Information Retrieval Techniques	PEC	3	0	0	3	3
4.	CP4092	Data Visualization Techniques	PEC	3	0	0	3	3
5.	BD4001	Principles of Supply Chain Management	PEC	3	0	0	3	3
6.	BD4002	Computational Geometry	PEC	3	0	0	3	3

SEMESTER III, ELECTIVE II

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	MP4093	Soft Computing Techniques	PEC	3	0	0	3	3
2.	MU4152	Multimedia Communication Networks	PEC	3	0	0	3	3
3.	BD4003	Parallel and Distributed Computing	PEC	3	0	0	3	3
4.	BD4091	Predictive Modeling	PEC	3	0	0	3	3
5.	BD4004	Image Processing and Analysis	PEC	3	0	0	3	3

SEMESTER III, ELECTIVE III

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	MP4091	Cognitive computing	PEC	3	0	0	3	3
2.	IF4095	Social Network analysis	PEC	3	0	0	3	3
3.	BD4005	Virtualization techniques and applications	PEC	3	0	0	3	3
4.	ML4291	Natural Language Processing	PEC	2	0	2	4	3

SEMESTER III, ELECTIVE IV

S. NO.	COURSE CODE	COURSE TITLE	CATE-GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	BD4006	Data Intensive Computing	PEC	3	0	0	3	3
2.	BD4007	R Language for Mining	PEC	3	0	0	3	3
3.	CP4097	Web analytics	PEC	3	0	0	3	3
4.	MP4071	Healthcare Analytics	PEC	3	0	0	3	3
5.	BD4008	Statistics for Business Analytics	PEC	3	0	0	3	3

AUDIT COURSES (AC)

Registration for any of these courses is optional to students

SL. NO	COURSE CODE	COURSE TITLE	PERIODS PER WEEK			CREDITS
			L	T	P	
1.	AX4091	English for Research Paper Writing	2	0	0	0
2.	AX4092	Disaster Management	2	0	0	0
3.	AX4093	Constitution of India	2	0	0	0
4.	AX4094	நற்றமிழ் இலக்கியம்	2	0	0	0

FOUNDATION COURSES (FC)

S. NO	COURSE CODE	COURSE TITLE	PERIODS PER WEEK			CREDITS	SEMESTER
			Lecture	Tutorial	Practical		
1.	MA4109	Applied Probability and Statistics	4	0	0	4	I

PROFESSIONAL CORE COURSES (PCC)

S. NO	COURSE CODE	COURSE TITLE	PERIODS PER WEEK			CREDITS	SEMESTER
			Lecture	Tutorial	Practical		
1.	CP4151	Advanced Data Structures and Algorithms	3	0	0	3	I
2.	BD4151	Foundations of Data Science	3	0	0	3	I
3.	CP4252	Machine Learning	3	0	2	4	I
4.	CP4152	Database Practices	3	0	2	4	I
5.	CP4161	Advanced Data Structures and Algorithms Laboratory	0	0	4	2	I
6.	BD4111	Big Data Computing Laboratory	0	0	2	1	I
7.	BD4201	Big Data Security	3	0	0	3	II
8.	BD4251	Big Data Mining and Analytics	3	0	0	3	II
9.	MP4251	Cloud Computing Technologies	3	0	0	3	II
10.	BD4202	Information Storage Management	3	0	0	3	II
11.	MP4151	Embedded Systems and IIOT	3	0	2	4	II
12.	BD4211	Big Data Mining And Analytics Laboratory	0	0	4	2	II

RESEARCH METHODOLOGY AND IPR COURSES (RMC)

S. NO	COURSE CODE	COURSE TITLE	PERIODS PER WEEK			CREDITS	SEMESTER
			Lecture	Tutorial	Practical		
1.	RM4151	Research Methodology and IPR	2	0	0	2	1

EMPLOYABILITY ENHANCEMENT COURSES (EEC)

S. NO	COURSE CODE	COURSE TITLE	PERIODS PER WEEK			CREDITS	SEMESTER
			Lecture	Tutorial	Practical		
1.	BD4211	Term Paper Writing and seminar	0	0	2	1	II
2.	BD 4311	Project Work I	0	0	12	6	III
3.	BD 4411	Project Work II	0	0	24	12	IV

SUMMARY

Sl. No.	NAME OF THE PROGRAMME: M.E. BIG DATA ANALYTICS					
	SUBJECT AREA	CREDITS PER SEMESTER				CREDITS TOTAL
		I	II	III	IV	
1.	FC	04	00	00	00	04
2.	PCC	17	18	00	00	35
3.	PEC	00	03	09	00	12
4.	RMC	02	00	00	00	02
5.	OEC	00	00	03	00	03
6.	EEC	00	01	06	12	19
7.	Non Credit/Audit Course	✓	✓	00	00	
8.	**TOTAL CREDIT**	**23**	**22**	**18**	**12**	**74**

MA4109 **APPLIED PROBABILITY AND STATISTICS** **L T P C**
4 0 0 4

COURSE OBJECTIVES:

- To compute probabilities and moments of standard distributions.
- To gain knowledge about regression and correlation.
- To provide the most appropriate estimator of the parameter in statistical inference.
- To decide whether to accept or reject specific values of a parameter.
- To understand many real-world problems fall naturally within the framework of multivariate normal theory.

UNIT I PROBABILITY AND RANDOM VARIABLES **12**

Probability – Axioms of probability – Conditional probability – Bayes theorem - Random variables - Probability function – Moments – Moment generating functions and their properties – Binomial, Poisson, Geometric, Uniform, Exponential, Gamma and Normal distributions – Function of a random variable.

UNIT II TWO DIMENSIONAL RANDOM VARIABLES **12**

Joint distributions – Marginal and conditional distributions – Functions of two dimensional random variables – Regression curve – Correlation.

UNIT III ESTIMATION THEORY **12**

Unbiased estimators – Method of moments – Maximum likelihood estimation - Curve fitting by principle of least squares – Regression lines.

UNIT IV TESTING OF HYPOTHESIS **12**

Sampling distributions – Type I and Type II errors – Small and large samples – Tests based on Normal, t, Chi square and F distributions for testing of mean, variance and proportions – Tests for independence of attributes and goodness of fit.

UNIT V MULTIVARIATE ANALYSIS **12**

Random vectors and matrices – Mean vectors and covariance matrices – Multivariate normal density and its properties – Principal components - Population principal components – Principal components from standardized variables

TOTAL : 60 PERIODS

COURSE OUTCOMES :

After completing this course, students should demonstrate competency in the following topics:

- Basic probability axioms and rules and the moments of discrete and continuous random variables.
- To deal with problems involving two dimensional random variables.
- Consistency, efficiency and unbiasedness of estimators, method of maximum likelihood estimation and Central Limit Theorem.
- Use statistical tests in testing hypotheses on data.
- Perform exploratory analysis of multivariate data, such as multivariate normal density, calculating descriptive statistics, testing for multivariate normality.

REFERENCES :

1. Devore, J. L., “Probability and Statistics for Engineering and the Sciences”, 8th Edition, Cengage Learning, 2014.

2. Dallas E. Johnson, “Applied Multivariate Methods for Data Analysis”, Thomson and Duxbury press, 1998.
3. Gupta S.C. and Kapoor V.K.,” Fundamentals of Mathematical Statistics”, 12th Edition, Sultan and Sons, New Delhi, 2020.
4. Johnson, R.A., Miller, I and Freund J., "Miller and Freund’s Probability and Statistics for Engineers ", 9th Edition, Pearson Education, Asia, 2016.
5. Richard A. Johnson and Dean W. Wichern, “Applied Multivariate Statistical Analysis”, 6th Edition, Pearson Education, Asia, 2012.

RM4151 RESEARCH METHODOLOGY AND IPR L T P C
2 0 0 2

UNIT I RESEARCH DESIGN 6

Overview of research process and design, Use of Secondary and exploratory data to answer the research question, Qualitative research, Observation studies, Experiments and Surveys.

UNIT II DATA COLLECTION AND SOURCES 6

Measurements, Measurement Scales, Questionnaires and Instruments, Sampling and methods. Data - Preparing, Exploring, examining and displaying.

UNIT III DATA ANALYSIS AND REPORTING 6

Overview of Multivariate analysis, Hypotheses testing and Measures of Association. Presenting Insights and findings using written reports and oral presentation.

UNIT IV INTELLECTUAL PROPERTY RIGHTS 6

Intellectual Property – The concept of IPR, Evolution and development of concept of IPR, IPR development process, Trade secrets, utility Models, IPR & Biodiversity, Role of WIPO and WTO in IPR establishments, Right of Property, Common rules of IPR practices, Types and Features of IPR Agreement, Trademark, Functions of UNESCO in IPR maintenance.

UNIT V PATENTS 6

Patents – objectives and benefits of patent, Concept, features of patent, Inventive step, Specification, Types of patent application, process E-filing, Examination of patent, Grant of patent, Revocation, Equitable Assignments, Licences, Licensing of related patents, patent agents, Registration of patent agents.

TOTAL : 30 PERIODS

REFERENCES:

1. Cooper Donald R, Schindler Pamela S and Sharma JK, “Business Research Methods”, Tata McGraw Hill Education, 11e (2012).
2. Catherine J. Holland, “Intellectual property: Patents, Trademarks, Copyrights, Trade Secrets”, Entrepreneur Press, 2007.
3. David Hunt, Long Nguyen, Matthew Rodgers, “Patent searching: tools & techniques”, Wiley, 2007.
4. The Institute of Company Secretaries of India, Statutory body under an Act of parliament, “Professional Programme Intellectual Property Rights, Law and practice”, September 2013.

CP4151 ADVANCED DATA STRUCTURES AND ALGORITHMS L T P C
3 0 0 3

COURSE OBJECTIVES:

- To understand the usage of algorithms in computing
- To learn and use hierarchical data structures and its operations
- To learn the usage of graphs and its applications
- To select and design data structures and algorithms that is appropriate for problems
- To study about NP Completeness of problems.

UNIT I ROLE OF ALGORITHMS IN COMPUTING & COMPLEXITY ANALYSIS 9

Algorithms – Algorithms as a Technology -Time and Space complexity of algorithms- Asymptotic analysis-Average and worst-case analysis-Asymptotic notation-Importance of efficient algorithms-Program performance measurement - Recurrences: The Substitution Method – The Recursion-Tree Method- Data structures and algorithms.

UNIT II HIERARCHICAL DATA STRUCTURES 9

Binary Search Trees: Basics – Querying a Binary search tree – Insertion and Deletion- Red Black trees: Properties of Red-Black Trees – Rotations – Insertion – Deletion -B-Trees: Definition of B - trees – Basic operations on B-Trees – Deleting a key from a B-Tree- Heap – Heap Implementation – Disjoint Sets - Fibonacci Heaps: structure – Mergeable-heap operations- Decreasing a key and deleting a node-Bounding the maximum degree.

UNIT III GRAPHS 9

Elementary Graph Algorithms: Representations of Graphs – Breadth-First Search – Depth-First Search – Topological Sort – Strongly Connected Components- Minimum Spanning Trees: Growing a Minimum Spanning Tree – Kruskal and Prim- Single-Source Shortest Paths: The Bellman-Ford algorithm – Single-Source Shortest paths in Directed Acyclic Graphs – Dijkstra‘s Algorithm; Dynamic Programming - All-Pairs Shortest Paths: Shortest Paths and Matrix Multiplication – The Floyd-Warshall Algorithm

UNIT IV ALGORITHM DESIGN TECHNIQUES 9

Dynamic Programming: Matrix-Chain Multiplication – Elements of Dynamic Programming – Longest Common Subsequence- Greedy Algorithms: – Elements of the Greedy Strategy- An Activity-Selection Problem - Huffman Coding.

UNIT V NP COMPLETE AND NP HARD 9

NP-Completeness: Polynomial Time – Polynomial-Time Verification – NP- Completeness and Reducibility – NP-Completeness Proofs – NP-Complete Problems.

TOTAL : 45 PERIODS

SUGGESTED ACTIVITIES:

1. Write an algorithm for Towers of Hanoi problem using recursion and analyze the complexity (No of disc-4)
2. Write any one real time application of hierarchical data structure
3. Write a program 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
4. Find the minimum cost to reach last cell of the matrix from its first cell
5. Discuss about any NP completeness problem

COURSE OUTCOMES:

CO1: Design data structures and algorithms to solve computing problems.

CO2: Choose and implement efficient data structures and apply them to solve problems.

CO3: Design algorithms using graph structure and various string-matching algorithms to solve real-life problems.

CO4: Design one's own algorithm for an unknown problem.

CO5: Apply suitable design strategy for problem solving.

REFERENCES:

1. S.Sridhar," Design and Analysis of Algorithms", Oxford University Press, 1st Edition, 2014.
2. Adam Drozdex, "Data Structures and algorithms in C++", Cengage Learning, 4th Edition, 2013.
3. T.H. Cormen, C.E.Leiserson, R.L. Rivest and C.Stein, "Introduction to Algorithms", Prentice Hall of India, 3rd Edition, 2012.
4. Mark Allen Weiss, "Data Structures and Algorithms in C++", Pearson Education, 3rd Edition, 2009.
5. E. Horowitz, S. Sahni and S. Rajasekaran, "Fundamentals of Computer Algorithms", University Press, 2nd Edition, 2008.
6. Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, "Data Structures and Algorithms", Pearson Education, Reprint 2006.

BD4151 FOUNDATIONS OF DATA SCIENCE L T P C

3 0 0 3

COURSE OBJECTIVES:

- To apply fundamental algorithms to process data.
- Learn to apply hypotheses and data into actionable predictions.
- Document and transfer the results and effectively communicate the findings using visualization techniques.
- To learn statistical methods and machine learning algorithms required for Data Science.
- To develop the fundamental knowledge and understand concepts to become a data science professional.

UNIT I INTRODUCTION TO DATA SCIENCE 9

Data science process – roles, stages in data science project – working with data from files – working with relational databases – exploring data – managing data – cleaning and sampling for modeling and validation – introduction to NoSQL.

UNIT II MODELING METHODS 9

Choosing and evaluating models – mapping problems to machine learning, evaluating clustering models, validating models – cluster analysis – K-means algorithm, Naïve Bayes – Memorization Methods – Linear and logistic regression – unsupervised methods.

UNIT III INTRODUCTION TO R 9

Reading and getting data into R – ordered and unordered factors – arrays and matrices – lists and data frames – reading data from files – probability distributions – statistical models in R -

manipulating objects – data distribution.

UNIT IV MAP REDUCE 9

Introduction – distributed file system – algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce – Hadoop - Understanding the Map Reduce architecture - Writing Hadoop MapReduce Programs - Loading data into HDFS - Executing the Map phase - Shuffling and sorting - Reducing phase execution.

UNIT V DATA VISUALIZATION 9

Documentation and deployment – producing effective presentations – Introduction to graphical analysis – plot() function – displaying multivariate data – matrix plots – multiple plots in one window - exporting graph using graphics parameters - Case studies.

TOTAL : 45 PERIODS

COURSE OUTCOMES:

CO1: Obtain, clean/process and transform data.

CO2: Analyze and interpret data using an ethically responsible approach.

CO3: Use appropriate models of analysis, assess the quality of input, derive insight from results, and investigate potential issues.

CO4: Apply computing theory, languages and algorithms, as well as mathematical and statistical models, and the principles of optimization to appropriately formulate and use data analyses.

CO5: Formulate and use appropriate models of data analysis to solve business-related challenges.

REFERENCES

1. Nina Zumel, John Mount, “Practical Data Science with R”, Manning Publications, 2014.
2. Mark Gardener, “Beginning R - The Statistical Programming Language”, John Wiley & Sons, Inc., 2012.
3. W. N. Venables, D. M. Smith and the R Core Team, “An Introduction to R”, 2013.
4. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta, “Practical Data Science Cookbook”, Packt Publishing Ltd., 2014.
5. Nathan Yau, “Visualize This: The FlowingData Guide to Design, Visualization, and Statistics”, Wiley, 2011.
6. Boris Lublinsky, Kevin T. Smith, Alexey Yakubovich, “Professional Hadoop Solutions”,John Wiley & Sons Inc., 2013.

CP4252 MACHINE LEARNING L T P C

3 0 2 4

COURSE OBJECTIVES:

- To understand the concepts and mathematical foundations of machine learning and types of problems tackled by machine learning
- To explore the different supervised learning techniques including ensemble methods
- To learn different aspects of unsupervised learning and reinforcement learning
- To learn the role of probabilistic methods for machine learning
- To understand the basic concepts of neural networks and deep learning

UNIT I INTRODUCTION AND MATHEMATICAL FOUNDATIONS 9

What is Machine Learning? Need –History – Definitions – Applications - Advantages, Disadvantages & Challenges -Types of Machine Learning Problems – Mathematical Foundations - Linear Algebra & Analytical Geometry -Probability and Statistics- Bayesian Conditional Probability -Vector Calculus & Optimization - Decision Theory - Information theory

UNIT II SUPERVISED LEARNING 9

Introduction-Discriminative and Generative Models -Linear Regression - Least Squares -Under-fitting / Overfitting -Cross-Validation – Lasso Regression- Classification - Logistic Regression- Gradient Linear Models -Support Vector Machines –Kernel Methods -Instance based Methods - K-Nearest Neighbors - Tree based Methods –Decision Trees –ID3 – CART - Ensemble Methods –Random Forest - Evaluation of Classification Algorithms

UNIT III UNSUPERVISED LEARNING AND REINFORCEMENT LEARNING 9

Introduction - Clustering Algorithms -K – Means – Hierarchical Clustering - Cluster Validity - Dimensionality Reduction –Principal Component Analysis – Recommendation Systems - EM algorithm. Reinforcement Learning – Elements -Model based Learning – Temporal Difference Learning

UNIT IV PROBABILISTIC METHODS FOR LEARNING 9

Introduction -Naïve Bayes Algorithm -Maximum Likelihood -Maximum Apriori -Bayesian Belief Networks -Probabilistic Modelling of Problems -Inference in Bayesian Belief Networks – Probability Density Estimation - Sequence Models – Markov Models – Hidden Markov Models

UNIT V NEURAL NETWORKS AND DEEP LEARNING 9

Neural Networks – Biological Motivation- Perceptron – Multi-layer Perceptron – Feed Forward Network – Back Propagation-Activation and Loss Functions- Limitations of Machine Learning – Deep Learning– Convolution Neural Networks – Recurrent Neural Networks – Use cases

45 PERIODS

SUGGESTED ACTIVITIES:

1. Give an example from our daily life for each type of machine learning problem
2. Study at least 3 Tools available for Machine Learning and discuss pros & cons of each
3. Take an example of a classification problem. Draw different decision trees for the example and explain the pros and cons of each decision variable at each level of the tree
4. Outline 10 machine learning applications in healthcare
5. Give 5 examples where sequential models are suitable.
6. Give at least 5 recent applications of CNN

PRACTICAL EXERCISES: 30 PERIODS

1. Implement a Linear Regression with a Real Dataset (https://www.kaggle.com/harrywang/housing). Experiment with different features in building a model. Tune the model's hyperparameters.
2. Implement a binary classification model. That is, answers a binary question such as "Are houses in this neighborhood above a certain price?"(use data from exercise 1). Modify the classification threshold and determine how that modification influences the model. Experiment with different classification metrics to determine your model's effectiveness.
3. Classification with Nearest Neighbors. In this question, you will use the scikit-learn's KNN classifier to classify real vs. fake news headlines. The aim of this question is for you to read

the scikit-learn API and get comfortable with training/validation splits. Use California Housing Dataset

4. In this exercise, you'll experiment with validation sets and test sets using the dataset. Split a training set into a smaller training set and a validation set. Analyze deltas between training set and validation set results. Test the trained model with a test set to determine whether your trained model is overfitting. Detect and fix a common training problem.
5. Implement the k-means algorithm using https://archive.ics.uci.edu/ml/datasets/Codon+usage dataset
6. Implement the Naïve Bayes Classifier using https://archive.ics.uci.edu/ml/datasets/Gait+Classification dataset
7. Project - (in Pairs) Your project must implement one or more machine learning algorithms and apply them to some data.

    a. Your project may be a comparison of several existing algorithms, or it may propose a new algorithm in which case you still must compare it to at least one other approach.
    b. You can either pick a project of your own design, or you can choose from the set of pre-defined projects.
    c. You are free to use any third-party ideas or code that you wish as long as it is publicly available.
    d. You must properly provide references to any work that is not your own in the write-up.
    e. Project proposal You must turn in a brief project proposal. Your project proposal should describe the idea behind your project. You should also briefly describe software you will need to write, and papers (2-3) you plan to read.

List of Projects (datasets available)

1. Sentiment Analysis of Product Reviews
2. Stock Prediction
3. Sales Forecasting
4. Music Recommendation
5. Handwriting Digit Classification
6. Fake News Detection
7. Sports Prediction
8. Object Detection
9. Disease Prediction

COURSE OUTCOMES:

Upon the completion of course, students will be able to

CO1: Understand and outline problems for each type of machine learning

CO2: Design a Decision tree and Random forest for an application

CO3: Implement Probabilistic Discriminative and Generative algorithms for an application and analyze the results.

CO4: Use a tool to implement typical Clustering algorithms for different types of applications.

CO5: Design and implement an HMM for a Sequence Model type of application and identify applications suitable for different types of Machine Learning with suitable justification.

TOTAL:75 PERIODS

REFERENCES

1. Stephen Marsland, “Machine Learning: An Algorithmic Perspective”, Chapman & Hall/CRC, 2nd Edition, 2014.

2. Kevin Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012
3. Ethem Alpaydin, “Introduction to Machine Learning”, Third Edition, Adaptive Computation and Machine Learning Series, MIT Press, 2014
4. Tom M Mitchell, “Machine Learning”, McGraw Hill Education, 2013.
5. Peter Flach, “Machine Learning: The Art and Science of Algorithms that Make Sense of Data”, First Edition, Cambridge University Press, 2012.
6. Shai Shalev-Shwartz and Shai Ben-David, “Understanding Machine Learning: From Theory to Algorithms”, Cambridge University Press, 2015
7. Christopher Bishop, “Pattern Recognition and Machine Learning”, Springer, 2007.
8. Hal Daumé III, “A Course in Machine Learning”, 2017 (freely available online)
9. Trevor Hastie, Robert Tibshirani, Jerome Friedman, “The Elements of Statistical Learning”, Springer, 2009 (freely available online)
10. Aurélien Géron , Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition, o'reilly, (2017)

CP4152 DATABASE PRACTICES L T P C
3 0 2 4

COURSE OBJECTIVES

- Describe the fundamental elements of relational database management systems
- Explain the basic concepts of relational data model, entity-relationship model, relational database design, relational algebra and SQL.
- Understand query processing in a distributed database system
- Understand the basics of XML and create well-formed and valid XML documents.
- Distinguish the different types of NoSQL databases
- To understand the different models involved in database security and their applications in real time world to protect the database and information associated with them.

UNIT I RELATIONAL DATA MODEL 15

Entity Relationship Model – Relational Data Model – Mapping Entity Relationship Model to Relational Model – Relational Algebra – Structured Query Language – Database Normalization.

Suggested Activities:

Data Definition Language

- Create, Alter and Drop
- Enforce Primary Key, Foreign Key, Check, Unique and Not Null Constraints
- Creating Views

Data Manipulation Language

- Insert, Delete, Update
- Cartesian Product, Equi Join, Left Outer Join, Right Outer Join and Full Outer Join
- Aggregate Functions
- Set Operations
- Nested Queries

Transaction Control Language

- Commit, Rollback and Save Points

UNIT II DISTRIBUTED DATABASES, ACTIVE DATABASES AND OPEN DATABASE CONNECTIVITY 15

Distributed Database Architecture – Distributed Data Storage – Distributed Transactions – Distributed Query Processing – Distributed Transaction Management – Event Condition Action Model – Design and Implementation Issues for Active Databases – Open Database Connectivity.

Suggested Activities:

- Distributed Database Design and Implementation
- Row Level and Statement Level Triggers
- Accessing a Relational Database using PHP, Python and R

UNIT III XML DATABASES 15

Structured, Semi structured, and Unstructured Data – XML Hierarchical Data Model – XML Documents – Document Type Definition – XML Schema – XML Documents and Databases – XML Querying – XPath – XQuery

Suggested Activities:

- Creating XML Documents, Document Type Definition and XML Schema
- Using a Relational Database to store the XML documents as text
- Using a Relational Database to store the XML documents as data elements
- Creating or publishing customized XML documents from pre-existing relational databases
- Extracting XML Documents from Relational Databases
- XML Querying

UNIT IV NOSQL DATABASES AND BIG DATA STORAGE SYSTEMS 15

NoSQL – Categories of NoSQL Systems – CAP Theorem – Document-Based NoSQL Systems and MongoDB – MongoDB Data Model – MongoDB Distributed Systems Characteristics – NoSQL Key-Value Stores – DynamoDB Overview – Voldemort Key-Value Distributed Data Store – Wide Column NoSQL Systems – Hbase Data Model – Hbase Crud Operations – Hbase Storage and Distributed System Concepts – NoSQL Graph Databases and Neo4j – Cypher Query Language of Neo4j – Big Data – MapReduce – Hadoop – YARN.

Suggested Activities:

- Creating Databases using MongoDB, DynamoDB, Voldemort Key-Value Distributed Data Store Hbase and Neo4j.
- Writing simple queries to access databases created using MongoDB, DynamoDB, Voldemort Key-Value Distributed Data Store Hbase and Neo4j.

UNIT V DATABASE SECURITY 15

Database Security Issues – Discretionary Access Control Based on Granting and Revoking Privileges – Mandatory Access Control and Role-Based Access Control for Multilevel Security – SQL Injection – Statistical Database Security – Flow Control – Encryption and Public Key Infrastructures – Preserving Data Privacy – Challenges to Maintaining Database Security – Database Survivability – Oracle Label-Based Security.

Suggested Activities:

Implementing Access Control in Relational Databases

TOTAL : 75 PERIODS

COURSE OUTCOMES

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

- Convert the ER-model to relational tables, populate relational database and formulate SQL queries on data.
- Understand and write well-formed XML documents
- Be able to apply methods and techniques for distributed query processing.
- Design and Implement secure database systems.
- Use the data control, definition, and manipulation languages of the NoSQL databases

REFERENCES:

1. R. Elmasri, S.B. Navathe, “Fundamentals of Database Systems”, Seventh Edition, Pearson Education 2016.
2. Henry F. Korth, Abraham Silberschatz, S. Sudharshan, “Database System Concepts”, Seventh Edition, McGraw Hill, 2019.
3. C.J.Date, A.Kannan, S.Swamynathan, “An Introduction to Database Systems, Eighth Edition, Pearson Education, 2006
4. Raghu Ramakrishnan , Johannes Gehrke “Database Management Systems”, Fourth Edition, McGraw Hill Education, 2015.
5. Harrison, Guy, “Next Generation Databases, NoSQL and Big Data” , First Edition, Apress publishers, 2015
6. Thomas Cannolly and Carolyn Begg, “Database Systems, A Practical Approach to Design, Implementation and Management”, Sixth Edition, Pearson Education, 2015

CP4161	ADVANCED DATA STRUCTURES AND ALGORITHMS LABORATORY	L T P C 0 0 4 2

COURSE OBJECTIVES:

- To acquire the knowledge of using advanced tree structures
- To learn the usage of heap structures
- To understand the usage of graph structures and spanning trees
- To understand the problems such as matrix chain multiplication, activity selection and Huffman coding
- To understand the necessary mathematical abstraction to solve problems.

LIST OF EXPERIMENTS:

1. Implementation of recursive function for tree traversal and Fibonacci
2. Implementation of iteration function for tree traversal and Fibonacci
3. Implementation of Merge Sort and Quick Sort
4. Implementation of a Binary Search Tree
5. Red-Black Tree Implementation
6. Heap Implementation
7. Fibonacci Heap Implementation
8. Graph Traversals
9. Spanning Tree Implementation
10. Shortest Path Algorithms (Dijkstra's algorithm, Bellman Ford Algorithm)
11. Implementation of Matrix Chain Multiplication
12. Activity Selection and Huffman Coding Implementation

HARDWARE/SOFTWARE REQUIREMENTS

1. 64-bit Open source Linux or its derivative
2. Open Source C++ Programming tool like G++/GCC

TOTAL : 60 PERIODS

COURSE OUTCOMES:

CO1: Design and implement basic and advanced data structures extensively
CO2: Design algorithms using graph structures
CO3: Design and develop efficient algorithms with minimum complexity using design techniques
CO4: Develop programs using various algorithms.
CO5: Choose appropriate data structures and algorithms, understand the ADT/libraries, and use it to design algorithms for a specific problem.

REFERENCES:

1. Lipschutz Seymour, “Data Structures Schaum's Outlines Series”, Tata McGraw Hill, 3rd Edition, 2014.
2. Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, “Data Structures and Algorithms”, Pearson Education, Reprint 2006.
3. http://www.coursera.org/specializations/data-structures-algorithms
4. http://www.tutorialspoint.com/data_structures_algorithms
5. http://www.geeksforgeeks.org/data-structures/

BD4111	BIG DATA COMPUTING LABORATORY	L T P C
		0 0 2 1

COURSE OBJECTIVES:

- To set up single and multi-node Hadoop Clusters.
- To solve Big Data problems using Map Reduce Technique.
- To learn NoSQL queries.
- To design algorithms that uses Map Reduce Technique to apply on Unstructured and structured data.
- To learn Scalable machine learning using Mahout.

LIST OF EXPERIMENTS:

1. Set up a pseudo-distributed, single-node Hadoop cluster backed by the Hadoop Distributed File System, running on Ubuntu Linux. After successful installation on one node, configuration of a multi-node Hadoop cluster (one master and multiple slaves).
2. MapReduce application for word counting on Hadoop cluster.
3. Unstructured data into NoSQL data and do all operations such as NoSQL query with API.
4. K-means clustering using map reduce.
5. Page Rank Computation.
6. Mahout machine learning library to facilitate the knowledge build up in big data analysis.
7. Application of Recommendation Systems using Hadoop/mahout libraries.

HARDWARE/SOFTWARE REQUIREMENTS

1. Java
2. Hadoop
3. Mahout
4. HBase/MongoDB

COURSE OUTCOMES:

CO1: Set up single and multi-node Hadoop Clusters.
CO2: Apply Map Reduce technique for various algorithms.
CO3: Design new algorithms that use Map Reduce to apply on Unstructured and structured data.
CO4: Develop Scalable machine learning algorithms for various Big data applications using Mahout.
CO5: Represent NoSQL data.

TOTAL: 30 PERIODS

REFERENCES:

1. Kristina Chodorow, “MongoDB: The Definitive Guide – Powerful and Scalable Data Storage”, O'Reilly, 3rd Edition, 2019.
2. Lars George, “HBase: The Definitive Guide”, O'Reilly, 2015.
3. Tom White, “Hadoop: The Definitive Guide – Storage and Analysis at Internet Scale”, O'Reilly, 4th Edition, 2015.
4. Robin Anil, Sean Owen, Ellen G. Friedman, Ted Dunning, “Mahout in Action”, Manning Publications, 2011.

BD4201 BIG DATA SECURITY L T P C
3 0 0 3

COURSE OBJECTIVES:

- To understand the mathematical foundations of security principles
- To appreciate the different aspects of encryption techniques
- To understand the role played by authentication in security
- To understand the security concerns of big-data.

UNIT I SYMMETRIC TECHNIQUES 9

Probability and Information Theory - Algebraic foundations – Number theory - Substitution Ciphers – Transposition Ciphers – Classical Ciphers – DES – AES – Confidentiality Modes of Operation

UNIT II ASYMMETRIC TECHNIQUES 9

Diffie-Hellman Key Exchange protocol – Discrete logarithm problem – RSA cryptosystems & cryptanalysis – ElGamal cryptosystem – Elliptic curve architecture and cryptography - Data Integrity techniques.

UNIT III AUTHENTICATION 9

Authentication requirements – Authentication functions – Message authentication codes – Hash functions – Security of hash functions and MACS – MD5 Message Digest algorithm – Secure hash algorithm.

UNIT IV SECURITY ANALYTICS I 9

Introduction to Security Analytics – Techniques in Analytics – Analysis in everyday life –Challenges in Intrusion and Incident Identification – Analysis of Log file – Simulation and Security Process.

UNIT V SECURITY ANALYTICS II **9**

Access Analytics – Security Analysis with Text Mining – Security Intelligence – Security Breaches.

TOTAL: 45 PERIODS

COURSE OUTCOMES:

After the completion of this course, students will be able to

CO1: Design algorithms in a secure manner for Big data applications

CO2: Use available security practices in big-data analytics.

CO3: Understand Mathematical foundations of security principles and different aspects of encryption techniques.

CO4: Explain the role played by authentication in security.

CO5: Analyze and find solutions for Security concerns of big-data.

REFERENCES

1. William Stallings, “Cryptography and Network security: Principles and Practices”,Pearson/PHI, 7th Edition, 2017.
2. Behrouz A. Forouzan, Debdeeep Mukhopadhyay “Cryptography and Network Security”, Tata McGraw Hill Education, 3rd Edition, 2015.
3. Douglas R. Stinson ,“Cryptography Theory and Practice ”, Chapman & Hall/CRC, 3rd Edition, 2021.
4. Mark Talabis, Robert McPherson, I Miyamoto and Jason Martin, “Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data”,Syngress Media, U.S.,2014

WEB REFERENCES:

1. *http://www.smartercomputingblog.com/category/big-data/*
2. *https://www.rd-alliance.org/group/big-data-ig-data-security-and-trust-wg/wiki/big-data-security-issues-challenges-tech-concerns*

ONLINE RESOURCES:

1. *https://www.tutorialspoint.com/big_data_tutorials.htm*
2. *https://www.simplilearn.com/tutorials/big-data-tutorial*

BD4251 BIG DATA MINING AND ANALYTICS **L T P C**

3 0 0 3

COURSE OBJECTIVES:

- To understand the computational approaches to Modeling, Feature Extraction
- To understand the need and application of Map Reduce
- To understand the various search algorithms applicable to Big Data
- To analyze and interpret streaming data
- To learn how to handle large data sets in main memory and learn the various clustering techniques applicable to Big Data

UNIT I DATA MINING AND LARGE SCALE FILES 9

Introduction to Statistical modeling – Machine Learning – Computational approaches to modeling – Summarization – Feature Extraction – Statistical Limits on Data Mining - Distributed File Systems – Map-reduce – Algorithms using Map Reduce – Efficiency of Cluster Computing Techniques.

UNIT II SIMILAR ITEMS 9

Nearest Neighbor Search – Shingling of Documents – Similarity preserving summaries – Locality sensitive hashing for documents – Distance Measures – Theory of Locality Sensitive Functions – LSH Families – Methods for High Degree of Similarities.

UNIT III MINING DATA STREAMS 9

Stream Data Model – Sampling Data in the Stream – Filtering Streams – Counting Distance Elements in a Stream – Estimating Moments – Counting Ones in Window – Decaying Windows.

UNIT IV LINK ANALYSIS AND FREQUENT ITEMSETS 9

Page Rank –Efficient Computation - Topic Sensitive Page Rank – Link Spam – Market Basket Model – A-priori algorithm – Handling Larger Datasets in Main Memory – Limited Pass Algorithm – Counting Frequent Item sets.

UNIT V CLUSTERING 9

Introduction to Clustering Techniques – Hierarchical Clustering –Algorithms – K-Means – CURE – Clustering in Non -– Euclidean Spaces – Streams and Parallelism – Case Study: Advertising on the Web – Recommendation Systems.

TOTAL: 45 PERIODS

COURSE OUTCOMES:

Upon completion of this course, the students will be able to

CO1: Design algorithms by employing Map Reduce technique for solving Big Data problems.
CO2: Design algorithms for Big Data by deciding on the apt Features set .
CO3: Design algorithms for handling petabytes of datasets
CO4: Design algorithms and propose solutions for Big Data by optimizing main memory consumption
CO5: Design solutions for problems in Big Data by suggesting appropriate clustering techniques.

REFERENCES:

1. Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, “Mining of Massive Datasets”, Cambridge University Press, 3rd Edition, 2020.
2. Jiawei Han, MichelineKamber, Jian Pei, “Data Mining Concepts and Techniques”, Morgan Kaufman Publications, Third Edition, 2012.
3. Ian H.Witten, Eibe Frank “Data Mining – Practical Machine Learning Tools and Techniques”, Morgan Kaufman Publications, Third Edition, 2011.
4. David Hand, HeikkiMannila and Padhraic Smyth, “Principles of Data Mining”, MIT PRESS, 2001

WEB REFERENCES:

1. *https://swayam.gov.in/nd2_arp19_ap60/preview*
2. *https://nptel.ac.in/content/storage2/nptel_data3/html/mhrd/ict/text/106104189/lec1.pdf*

ONLINE RESOURCES:

1. *https://examupdates.in/big-data-analytics/*
2. *https://www.tutorialspoint.com/big_data_analytics/index.htm*
3. *https://www.tutorialspoint.com/data_mining/index.htm*

MP4251 **CLOUD COMPUTING TECHNOLOGIES** **L T P C**
3 0 0 3

COURSE OBJECTIVES:

- To gain expertise in Virtualization, Virtual Machines and deploy practical virtualization solution
- To understand the architecture, infrastructure and delivery models of cloud computing.
- To explore the roster of AWS services and illustrate the way to make applications in AWS
- To gain knowledge in the working of Windows Azure and Storage services offered by Windows Azure
- To develop the cloud application using various programming model of Hadoop and Aneka

UNIT I VIRTUALIZATION AND VIRTUALIZATION INFRASTRUCTURE 6

Basics of Virtual Machines - Process Virtual Machines – System Virtual Machines –Emulation – Interpretation – Binary Translation - Taxonomy of Virtual Machines. Virtualization –Management Virtualization –— Hardware Maximization – Architectures – Virtualization Management – Storage Virtualization – Network Virtualization- Implementation levels of virtualization – virtualization structure – virtualization of CPU, Memory and I/O devices – virtual clusters and Resource Management – Virtualization for data center automation

UNIT II CLOUD PLATFORM ARCHITECTURE 12

Cloud Computing: Definition, Characteristics - Cloud deployment models: public, private, hybrid, community – Categories of cloud computing: Everything as a service: Infrastructure, platform, software- A Generic Cloud Architecture Design – Layered cloud Architectural Development – Architectural Design Challenges

UNIT III AWS CLOUD PLATFORM - IAAS 9

Amazon Web Services: AWS Infrastructure- AWS API- AWS Management Console - Setting up AWS Storage - Stretching out with Elastic Compute Cloud - Elastic Container Service for Kubernetes- AWS Developer Tools: AWS Code Commit, AWS Code Build, AWS Code Deploy, AWS Code Pipeline, AWS code Star - AWS Management Tools: Cloud Watch, AWS Auto Scaling, AWS control Tower, Cloud Formation, Cloud Trail, AWS License Manager

UNIT IV PAAS CLOUD PLATFORM 9

Windows Azure: Origin of Windows Azure, Features, The Fabric Controller – First Cloud APP in Windows Azure- Service Model and Managing Services: Definition and Configuration, Service runtime API- Windows Azure Developer Portal- Service Management API- Windows Azure Storage Characteristics-Storage Services- REST API- Blops

UNIT V PROGRAMMING MODEL 9

Introduction to Hadoop Framework - Mapreduce, Input splitting, map and reduce functions, specifying input and output parameters, configuring and running a job –Developing Map Reduce Applications - Design of Hadoop file system –Setting up Hadoop Cluster- Aneka: Cloud Application Platform, Thread Programming, Task Programming and Map-Reduce Programming in Aneka

COURSE OUTCOMES:

CO1: Employ the concepts of virtualization in the cloud computing
CO2: Identify the architecture, infrastructure and delivery models of cloud computing
CO3: Develop the Cloud Application in AWS platform
CO4: Apply the concepts of Windows Azure to design Cloud Application
CO5: Develop services using various Cloud computing programming models.

TOTAL: 45 PERIODS

REFERENCES

1. Bernard Golden, Amazon Web Service for Dummies, John Wiley & Sons, 2013.
2. Raoul Alongi, AWS: The Most Complete Guide to Amazon Web Service from Beginner to Advanced Level, Amazon Asia- Pacific Holdings Private Limited, 2019.
3. Sriram Krishnan, Programming: Windows Azure, O'Reilly,2010.
4. Rajkumar Buyya, Christian Vacchiola, S.Thamarai Selvi, Mastering Cloud Computing , MCGraw Hill Education (India) Pvt. Ltd., 2013.
5. Danielle Ruest, Nelson Ruest, —Virtualization: A Beginner"s Guide‖, McGraw-Hill Osborne Media, 2009.
6. Jim Smith, Ravi Nair , "Virtual Machines: Versatile Platforms for Systems and Processes", Elsevier/Morgan Kaufmann, 2005.
7. John W.Rittinghouse and James F.Ransome, "Cloud Computing: Implementation, Management, and Security", CRC Press, 2010.
8. Toby Velte, Anthony Velte, Robert Elsenpeter, "Cloud Computing, A Practical Approach", McGraw-Hill Osborne Media, 2009.
9. Tom White, "Hadoop: The Definitive Guide", Yahoo Press, 2012.

BD4202 INFORMATION STORAGE MANAGEMENT L T P C
3 0 0 3

COURSE OBJECTIVES:

- To understand the storage architecture and technologies in Information management
- To learn to establish and manage a data center
- To learn various storage technologies for the required application
- To apply security measures to the data center

UNIT I STORAGE TECHNOLOGY 9

Review data creation - Amount of data being created - Understand the value of data to a business - Challenges in data storage and data management - Solutions available for data storage - Core elements of a data center infrastructure - Role of each element in supporting business activities.

UNIT II STORAGE SYSTEM ARCHITECTURE 9

Hardware and software components of the host environment - Key protocols and concepts used by each component - Physical and logical components of a connectivity environment - Major physical components of a disk drive and their function - Logical constructs of a physical disk - Access characteristics - Performance Implications - Concept of RAID and its components - Different RAID levels and their suitability for different application environments - Compare and contrast integrated and modular storage systems - High-level architecture and working of an intelligent storage system.

UNIT III INTRODUCTION TO NETWORKED STORAGE 9

Evolution of networked storage - Architecture - Components - Topologies of FC-SAN - NAS - IP-SAN - Benefits of the different networked storage options - Understand the need for long-Term archiving solutions - Describe how CAS fulfill the need - Understand the appropriateness - Different networked storage options - Different application environments.

UNIT IV INFORMATION AVAILABILITY, MONITORING & MANAGING DATA CENTERS 9

List reasons for planned or unplanned outages - Impact of downtime - Business continuity (BC) - Disaster recovery (DR) - RTO - RPO - Identify single points of failure - List solutions to mitigate failures - Architecture of backup/recovery - Different backup or recovery topologies - Replication technologies - Role in ensuring information availability and business continuity - Remote replication technologies - Role in providing disaster recovery and business continuity capabilities - Identify key areas to monitor in a data center - Industry standards for data center monitoring and management - Key metrics - Key management tasks.

UNIT V SECURING STORAGE AND STORAGE VIRTUALIZATION 9

Information security - Critical security attributes - Storage security domains - List and analyze the common threats in each domain - Virtualization technologies - Block-level and file-level virtualization technologies and processes**.**

TOTAL: 45 PERIODS

COURSE OUTCOMES:

Upon completion of the course, the student should be able to:

CO1: Understand the basics of storage management for Information maintenance.
CO2: Study the requirements and strategies for the data center.
CO3: Learn various storage technologies for the required application.
CO4: Apply security measures to the data center.
CO5: Analyze Quality of Service in Storage.

REFERENCES:

1. EMC Corporation, "Information Storage and Management: Storing, Managing, and Protecting Digital Information", 2nd Edition, Wiley, India, 2012.
2. Marc Farley, “Building Storage Networks”, Tata McGraw Hill” ,Osborne, 2001.
3. Ulf Troppens, Rainer Erkens, Wolfgang Muller-Friedt, Rainer Wolafka, Nils Haustein, “Storage Networks Explained: Basics and Application of Fibre Channel SAN, NAS, ISCSI, InfiniBand and FCoE”, Wiley, 2015.
4. Robert Spalding, “Storage Networks: The Complete Reference”, Tata McGraw Hill , 2017.

WEB REFERENCES:

1. https://nptel.ac.in/courses/106108058/.

ONLINE RESOURCES:

2. https://dokumen.tips/engineering/cp7029-information-storage-management-notes-58f9ada4e0e17.html.

MP4151 **EMBEDDED SYSTEMS AND IIOT** **L T P C**
3 0 2 4

COURSE OBJECTIVES:

- To learn the internal architecture of an embedded processor including timers and interrupts.
- To learn and use embedded C programming.
- To provide exposure on architecture and components of IIOT.
- To introduce the communication protocols of IIOT.
- To study about visualization and data processing of IIOT.

UNIT I EMBEDDED PROCESSOR 9

Embedded processors –8051 Microcontroller – Architecture, Instruction set and programming. Programming parallel ports, Timers and serial port – Memory and I/O devices interfacing – Interrupt handling.

UNIT II EMBEDDED C PROGRAMMING 9

Programming Embedded Systems in C - Memory And I/O Devices Interfacing - Implementing Timers, Interrupts and Serial communication in embedded C- Need For RTOS - Multiple Tasks and Processes – Context Switching - Priority Based Scheduling Policies.

UNIT III INTRODUCTION & ARCHITECTURE OF IIOT 9

Introduction to IOT, IIOT, IOT Vs. IIOT, Architecture of IIoT, IOT node- Components of IIOT - Fundamentals of Control System, introductions, components, closed loop & open loop system, IIOT System components: Sensors, Gateways, Routers, Modem, Cloud brokers, servers and its integration, Introduction to sensors, Types of sensors, working principle of basic Sensors - Ultrasonic Sensor, IR sensor, MQ2, Temperature and Humidity Sensors (DHT-11). Digital switch, Electro Mechanical switches, Roles of sensors and actuators in IIOT, Special requirements for IIOT sensors.

UNIT IV COMMUNICATION TECHNOLOGIES OF IIOT 9

Hardwire the sensors with different protocols such as HART, MODBUS-Serial & Parallel, Ethernet, BACNet , Current, M2M etc. Need of protocols; Communication Protocols: Wi-Fi, Wi-Fi direct, IEEE 802.15.4, Zigbee, Z wave, BLE, SPI, RFID, Industry standards communication technology (COAP, LoRAWAN, OPC UA, MQTT AMQP IIOT), connecting into existing Modbus and Profibus technology, wireless network communication.

UNIT V VISUALIZATION OF IIOT 9

Cloud platforms: Overview of cots cloud platforms, predix, thingworx, azure etc. Front-end EDGE devices, Enterprise data for IIoT, Emerging descriptive data standards for IIoT, Cloud database, Cloud computing, Fog or Edge computing. Connecting an Arduino/Raspberry pi to the Web: Introduction, setting up the Arduino/Raspberry pi development environment, Options for Internet connectivity with Arduino, Configuring your Arduino/Raspberry pi board for the IoT.

TOTAL: 45 PERIODS

SUGGESTED ACTIVITIES:

1. A Study on the various embedded processors like virtual watches, PDAS, digital cameras, mp3 players
2. Develop an application using embedded C programming in arduino
3. Build a project using IIOT components
4. Study of communication protocols and technology in IIOT
5. Presentation on most prominent IIOT visualization tools

PRACTICAL EXERCISES:

1. Experiments on Arduino, ESp8266, raspberry Pi
2. Measurement of temperature & pressure values of the process using raspberry pi/node mcu.
3. Modules and Sensors Interfacing (IR sensor, Ultrasonic sensors, Soil moisture sensor) using Raspberry pi/node mcu.
4. Modules and Actuators Interfacing (Relay, Motor, Buzzer) using Raspberry pi/node mcu.
5. Demonstration of MQTT communication
6. Demonstration of LoRa communication.

TOTAL: 30 PERIODS

HARDWARE/SOFTWARE REQUIREMENTS

1. Arduino
2. ESp8266
3. Raspberry Pi

COURSE OUTCOMES:

CO1: Describe the internal architecture of an embedded processor including timers and interrupts.
CO2: Write the embedded C programming.
CO3: Use the components of IIOT for building applications.
CO4: Demonstrate and perform the communication by using the protocols.
CO5: Explain about visualization and data processing of IIOT.

TOTAL: 75 PERIODS

REFERENCES:

1. Michael J. Pont, "Embedded C", Pearson Education, 2007.
2. Muhammed Ali Mazidi, Janice Gillispie Mazidi, Rolin D. McKinlay, "The 8051 Microcontroller and Embedded Systems", Pearson Education, Second Edition, 2014.
3. Mahmood, Zaigham(Ed), “The Internet of Things in the Industrial Sector”, Springer Publication, 2019.
4. Sabina Jeschke, Christian Brecher, Houbing Song, Danda B. Rawat, “Industrial Internet of Things: Cyber manufacturing System (wireless Technology)”, Springer Publication, 2017.
5. Hakima Chaouchi, “The Internet of Things Connecting Objects to the Web” ISBN: 978-1-84821-140-7, Willy Publications, 2010.
6. Robert Barton, Patrick Grossetete, David Hanes, Jerome Henry, Gonzalo Salgueiro, "loT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things", CISCO Press, 2017.
7. Ismail Butun, “Industrial IoT Challenges, Design Principles, Applications, and Security”, Springer Publications, 2020.
8. Wayne Wolf, "Computers as Components: Principles of Embedded Computer System Design", Elsevier, 2006.
9. David Etter, “IOT (Internet of Things) Programming: A Simple and Fast Way of Learning, IOT”, 2016.
10. Olivier Hersent, David Boswarthick, Omar Elloumi, The Internet of Things: Key Applications and Protocols, ISBN: 978-1-119-99435-0, 2 nd Edition, Willy Publications.
11. Daniel Minoli, “Building the Internet of Things with IPv6 and MIPv6: The Evolving World of M2M Communications”, ISBN: 978-1-118-47347-4, Willy Publications.

BD4211 **BIG DATA MINING AND ANALYTICS LABORATORY** **L T P C**
0 0 4 2

COURSE OBJECTIVES:

- To learn to process big data using Hadoop framework and MapReduce.
- To analyze big data using classification and clustering techniques.
- To realize storage of big data using MongoDB and Hbase.
- To develop big data applications for streaming data using Apache Spark.

LIST OF EXPERIMENTS:

1. Install, configure and run Hadoop and HDFS.
2. Implement word count / frequency programs using MapReduce(MR).
3. Implement an MR program that processes a weather dataset.
4. Implement SVM and clustering techniques using R.
5. Visualize data using any plotting framework.
6. Implement an application that stores big data in Hbase / MongoDB using Hadoop / R.
7. Install, deploy and configure Apache Spark cluster. Run an application using Apache Spark.

TOTAL: 60 PERIODS

LAB REQUIREMENTS FOR A BATCH OF 30 STUDENTS

SOFTWARE
Hadoop, R Package, Hbase, MongoDB, Apache Spark

COURSE OUTCOMES:

Upon completion of this course, the students will be able to

CO1: Process big data using Hadoop framework.

CO2: Implement MapReduce framework for processing big data.

CO3: Perform data analysis using classification and clustering techniques.

CO4: Realize storage of big data using MongoDB , Hbase and Apache Spark

CO5: Perform graphical data analysis

BD4212 **TERM PAPER WRITING AND SEMINAR** **L T P C**
0 0 2 1

In this course, students will develop their scientific and technical reading and writing skills that they need to understand and construct research articles. A term paper requires a student to obtain information from a variety of sources (i.e., Journals, dictionaries, reference books) and then place it in logically developed ideas. The work involves the following steps:

1. Selecting a subject, narrowing the subject into a topic
2. Stating an objective.
3. Collecting the relevant bibliography (atleast 15 journal papers)
4. Preparing a working outline.
5. Studying the papers and understanding the authors contributions and critically analysing each paper.
6. Preparing a working outline

7. Linking the papers and preparing a draft of the paper.
8. Preparing conclusions based on the reading of all the papers.
9. Writing the Final Paper and giving final Presentation

Please keep a file where the work carried out by you is maintained.
Activities to be carried out

Activity	Instructions	Submission week	Evaluation
Selection of area of interest and Topic Stating an Objective	You are requested to select an area of interest, topic and state an objective	2nd week	**3 %** Based on clarity of thought, current relevance and clarity in writing
Collecting Information about your area & topic	1. List 1 Special Interest Groups or professional society 2. List 2 journals 3. List 2 conferences, symposia or workshops 4. List 1 thesis title 5. List 3 web presences (mailing lists, forums, news sites) 6. List 3 authors who publish regularly in your area 7. Attach a call for papers (CFP) from your area.	3rd week	**3%** (the selected information must be area specific and of international and national standard)
Collection of Journal papers in the topic in the context of the objective – collect 20 & then filter	• You have to provide a complete list of references you will be using- Based on your objective - Search various digital libraries and Google Scholar • When picking papers to read - try to: • Pick papers that are related to each other in some ways and/or that are in the same field so that you can write a meaningful survey out of them, • Favour papers from well-known journals and conferences, • Favour "first" or "foundational" papers in the field (as indicated in other people's survey paper), • Favour more recent papers, • Pick a recent survey of the field so you can quickly gain an overview, • Find relationships with respect	4th week	**6%** (the list of standard papers and reason for selection)

	to each other and to your topic area (classification scheme/categorization) • Mark in the hard copy of papers whether complete work or section/sections of the paper are being considered		
Reading and notes for first 5 papers	Reading Paper Process • For each paper form a Table answering the following questions: • What is the main topic of the article? • What was/were the main issue(s) the author said they want to discuss? • Why did the author claim it was important? • How does the work build on other's work, in the author's opinion? • What simplifying assumptions does the author claim to be making? • What did the author do? • How did the author claim they were going to evaluate their work and compare it to others? • What did the author say were the limitations of their research? • What did the author say were the important directions for future research? Conclude with limitations/issues not addressed by the paper (from the perspective of your survey)	5th week	**8%** (the table given should indicate your understanding of the paper and the evaluation is based on your conclusions about each paper)
Reading and notes for next5 papers	Repeat Reading Paper Process	6th week	**8%** (the table given should indicate your understanding of the paper and the evaluation is based on your conclusions about each paper)

Reading and notes for final 5 papers	Repeat Reading Paper Process	7th week	**8%** (the table given should indicate your understanding of the paper and the evaluation is based on your conclusions about each paper)
Draft outline 1 and Linking papers	Prepare a draft Outline, your survey goals, along with a classification / categorization diagram	8th week	**8%** (this component will be evaluated based on the linking and classification among the papers)
Abstract	Prepare a draft abstract and give a presentation	9th week	**6%** (Clarity, purpose and conclusion) **6%** Presentation & Viva Voce
Introduction Background	Write an introduction and background sections	10th week	**5%** (clarity)
Sections of the paper	Write the sections of your paper based on the classification / categorization diagram in keeping with the goals of your survey	11thweek	**10%** (this component will be evaluated based on the linking and classification among the papers)
Your conclusions	Write your conclusions and future work	12th week	**5%** (conclusions – clarity and your ideas)
Final Draft	Complete the final draft of your paper	13th week	**10%** (formatting, English, Clarity and linking) **4%** Plagiarism

			Check Report
Seminar	A brief 15 slides on your paper	14th & 15th week	**10%** (based on presentation and Viva-voce)

TOTAL: 30 PERIODS

BD4071 HIGH PERFORMANCE COMPUTING FOR BIG DATA L T P C
3 0 0 3

COURSE OBJECTIVES:

- To learn the fundamental concepts of High Performance Computing.
- To learn the network & software infrastructure for high performance computing.
- To understand real time analytics using high performance computing.
- To learn the different ways of security perspectives and technologies used in HPC.
- To understand the emerging big data applications.

UNIT I INTRODUCTION 9

The Emerging IT Trends- IOT/IOE-Apache Hadoop for big data analytics-Big data into big insights and actions – Emergence of BDA discipline – strategic implications of big data – BDA Challenges – HPC paradigms – Cluster computing – Grid Computing – Cloud computing – Heterogeneous computing – Mainframes for HPC - Supercomputing for BDA – Appliances for BDA.

UNIT II NETWORK & SOFTWARE INFRASTRUCTURE FOR HIGH PERFORMANCE BDA 9

Design of Network Infrastructure for high performance BDA – Network Virtualization – Software Defined Networking – Network Functions Virtualization – WAN optimization for transfer of big data – started with SANs- storage infrastructure requirements for storing big data – FC SAN – IP SAN – NAS – GFS – Panasas – Luster file system – Introduction to cloud storage.

UNIT II REAL TIME ANALYTICS USING HIGH PERFORMANCE COMPUTING 9

Technologies that support Real time analytics – MOA: Massive online analysis – GPFS: General parallel file system – Client case studies – Key distinctions – Machine data analytics – operational analytics – HPC Architecture models – In Database analytics – In memory analytics

UNIT IV SECURITY AND TECHNOLOGIES 9

Security, Privacy and Trust for user – generated content: The challenges and solutions – Role of real time big data processing in the IoT – End to End Security Framework for big sensing data streams – Clustering in big data.

UNIT V EMERGING BIG DATA APPLICATIONS **9**

Deep learning Accelerators – Accelerators for clustering applications in machine learning - Accelerators for classification algorithms in machine learning – Accelerators for Big data Genome Sequencing

TOTAL: 45 PERIODS

COURSE OUTCOMES:

Upon completion of the course, the student should be able to:

CO1: Understand the basics concepts of High Performance computing systems.

CO2: Apply the concepts of network and software infrastructure for high performance computing

CO3: Use real time analytics using high performance computing.

CO4: Apply the security models and big data applications in high performance computing

CO5: Understand the emerging big data applications.

REFERENCES:

1. Pethuru Raj, Anupama Raman, Dhivya Nagaraj and Siddhartha Duggirala, "High-Performance Big-Data Analytics: Computing Systems and Approaches", Springer, 1st Edition, 2015.
2. "Big Data Management and Processing", Kuan-Ching Li , Hai Jiang, Albert Y. Zomaya, CRC Press,1st Edition,2017.
3. "High Performance Computing for Big Data: Methodologies and Applications", Chao wang ,CRC Press,1st Edition,2018
4. "High-Performance Data Mining And Big Data Analytics" , Khosrow Hassibi, Create Space Independent Publishing Platform,!st Edition,2014
5. "High performance computing: Modern systems and practices", Thomas Sterling, Matthew Anderson, Morgan Kaufmann publishers,1st Edition,2017

WEB REFERENCES:

1. https://www.hpcwire.com/

ONLINE RESOURCES:

1. *http://hpc.fs.uni-lj.si/sites/default/files/HPC_for_dummies.pdf*
2. *https://www.nics.tennessee.edu/computing-resources/what-is-hpc*

MP4094 WEB SERVICES AND API DESIGN **L T P C 3 0 0 3**

COURSE OBJECTIVES:

- To learn the basics of Web service.
- To become familiar with the Web Services building blocks
- To learn to work with RESTful web services.
- To implement the RESTful web services.
- To understand resource oriented Architecture

UNIT I INTRODUCTION TO WEB SERVICE **9**

Overview – Web service-Architecture – Service-Oriented Architecture (SOA), Architecting Web Services: Web Services Technology Stack, Logical Architectural View, Deployment Architectural View, and Process Architectural View.

UNIT II WEB SERVICE BUILDING BLOCKS 9

Introduction to SOAP: SOAP Syntax- Sending SOAP Messages - SOAP Implementations - Introduction to WSDL: WSDL Syntax - SOAP Binding - WSDL Implementations - Introduction to UDDI: The UDDI API - Implementations - The Future of UDDI

UNIT III RESTFUL WEB SERVICES 9

Programmable Web - HTTP: Documents in Envelopes - Method Information - Scoping Information - The Competing Architectures - Technologies on the Programmable Web -Leftover Terminology - Writing Web Service Clients: The Sample Application - Making the Request: HTTP Libraries - Processing the Response: XML Parsers - JSON Parsers: Handling Serialized Data - Clients Made Easy with WADL.

UNIT IV IMPLEMENTATION OF RESTFUL WEB SERVICES 9

Introducing the Simple Storage Service - Object-Oriented Design of S3 - Resources - HTTP Response Codes Resource- URIs - Addressability - Statelessness - Representations - Links and Connectedness - The Uniform Interface – Spring Web Services – Spring MVC Components - Spring Web Flow - A Service Implementation using Spring Data REST.

UNIT V RESOURCE ORIENTED ARCHITECTURE 9

Resource- URIs - Addressability - Statelessness - Representations - Links and Connectedness - The Uniform Interface- Designing Read-Only Resource-Oriented Services : Resource Design - Turning Requirements Into Read-Only Resources - Figure Out the Data Set- Split the Data Set into Resources- Name the Resources - Design Representation- Link the Resources to Each Other- The HTTP Response

COURSE OUTCOMES:

CO1: Explain how to write XML documents.
CO2: Apply the web service building blocks such as SOAP, WSDL and UDDI
CO3: Describe the RESTful web services.
CO4: Implement the RESTful web service with Spring Boot MVC
CO5: Discuss Resource-oriented Architecture.

TOTAL: 45 PERIODS

REFERENCES

1. Leonard Richardson and Sam Ruby, RESTful Web Services, O'Reilly Media, 2007
2. McGovern, et al., "Java Web Services Architecture", Morgan Kaufmann Publishers,2005.
3. Lindsay Bassett, Introduction to JavaScript Object Notation, O'Reilly Media, 2015
4. Craig Walls, "Spring in Action, Fifth Edition", Manning Publications, 2018
5. Raja CSP Raman, Ludovic Dewailly, "Building A RESTful Web Service with Spring 5", Packt Publishing, 2018 .
6. Bogunuva Mohanram Balachandar, "Restful Java Web Services, Third Edition: A pragmatic guide to designing and building RESTful APIs using Java", Ingram short title, 3rd Edition, 2017.
7. Mario-Leander Reimer, "Building RESTful Web Services with Java EE 8: Create modern RESTful web services with the Java EE 8 API", Packt publishing, 2018.

CP4093 **INFORMATION RETRIEVAL TECHNIQUES** **L T P C**
3 0 0 3

COURSE OBJECTIVES:

- To understand the basics of information retrieval with pertinence to modeling, query operations and indexing
- To get an understanding of machine learning techniques for text classification and clustering.
- To understand the various applications of information retrieval giving emphasis to multimedia IR, web search
- To get an understanding of machine learning techniques for text classification and clustering.
- To understand the concepts of digital libraries

UNIT I INTRODUCTION: MOTIVATION **9**

Basic Concepts – Practical Issues - Retrieval Process – Architecture - Boolean Retrieval –Retrieval Evaluation – Open-Source IR Systems–History of Web Search – Web Characteristics–The impact of the web on IR ––IR Versus Web Search–Components of a Search engine.

UNIT II MODELING **9**

Taxonomy and Characterization of IR Models – Boolean Model – Vector Model - Term Weighting – Scoring and Ranking –Language Models – Set Theoretic Models - Probabilistic Models – Algebraic Models – Structured Text Retrieval Models – Models for Browsing

UNIT III INDEXING **9**

Static and Dynamic Inverted Indices – Index Construction and Index Compression. Searching - Sequential Searching and Pattern Matching. Query Operations -Query Languages – Query Processing - Relevance Feedback and Query Expansion - Automatic Local and Global Analysis – Measuring Effectiveness and Efficiency

UNIT IV EVALUATION AND PARALLEL INFORMATION RETRIEVAL **9**

Traditional Effectiveness Measures – Statistics in Evaluation – Minimizing Adjudication Effect – Nontraditional Effectiveness Measures – Measuring Efficiency – Efficiency Criteria –Queueing Theory – Query Scheduling – Parallel Information Retrieval – Parallel Query Processing – MapReduce

UNIT V SEARCHING THE WEB **9**

Searching the Web –Structure of the Web –IR and web search – Static and Dynamic Ranking – Web Crawling and Indexing – Link Analysis - XML Retrieval Multimedia IR: Models and Languages – Indexing and Searching Parallel and Distributed IR – Digital Libraries.

COURSE OUTCOMES:

CO1: Build an Information Retrieval system using the available tools.
CO2: Identify and design the various components of an Information Retrieval system.
CO3: Categorize the different types of IR Models.
CO4: Apply machine learning techniques to text classification and clustering which is used for efficient Information Retrieval.
CO5: Design an efficient search engine and analyze the Web content structure.

TOTAL: 45 PERIODS

REFERENCES

1. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze, “Introduction to Information Retrieval, Cambridge University Press, First South Asian Edition, 2008.
2. Stefan Buttcher, Implementing and Evaluating Search Engines, The MIT Press, Cambridge, Massachusetts London, England, 2016.
3. Ricardo Baeza – Yates, Berthier Ribeiro – Neto, “Modern Information Retrieval: The concepts and Technology behind Search (ACM Press Books), Second Edition, 2011.
4. Stefan Buttcher, Charles L. A. Clarke, Gordon V. Cormack, “Information Retrieval

CP4092 DATA VISUALIZATION TECHNIQUES L T P C

3 0 0 3

COURSE OBJECTIVES:

- To develop skills to both design and critique visualizations.
- To introduce visual perception and core skills for visual analysis.
- To understand technological advancements of data visualization
- To understand various data visualization techniques
- To understand the methodologies used to visualize large data sets

UNIT I INTRODUCTION AND DATA FOUNDATION 9

Basics - Relationship between Visualization and Other Fields -The Visualization Process - Pseudo code Conventions - The Scatter plot. Data Foundation - Types of Data - Structure within and between Records - Data Preprocessing - Data Sets

UNIT II FOUNDATIONS FOR VISUALIZATION 9

Visualization stages - Semiology of Graphical Symbols - The Eight Visual Variables – Historical Perspective - Taxonomies - Experimental Semiotics based on Perception Gibson‘s Affordance theory – A Model of Perceptual Processing.

UNIT III VISUALIZATION TECHNIQUES 9

Spatial Data: One-Dimensional Data - Two-Dimensional Data – Three Dimensional Data - Dynamic Data - Combining Techniques. Geospatial Data : Visualizing Spatial Data - Visualization of Point Data -Visualization of Line Data - Visualization of Area Data – Other Issues in Geospatial Data Visualization Multivariate Data : Point-Based Techniques - LineBased Techniques - Region-Based Techniques - Combinations of Techniques – Trees Displaying Hierarchical Structures – Graphics and Networks- Displaying Arbitrary Graphs/Networks.

UNIT IV INTERACTION CONCEPTS AND TECHNIQUES 9

Text and Document Visualization: Introduction - Levels of Text Representations - The Vector Space Model - Single Document Visualizations -Document Collection Visualizations – Extended Text Visualizations Interaction Concepts: Interaction Operators - Interaction Operands and Spaces - A Unified Framework. Interaction Techniques: Screen Space - Object-Space –Data Space - Attribute Space- Data Structure Space - Visualization Structure – Animating Transformations - Interaction Control.

UNIT V RESEARCH DIRECTIONS IN VISUALIZATIONS 9

Steps in designing Visualizations – Problems in designing effective Visualizations- Issues of Data. Issues of Cognition, Perception, and Reasoning. Issues of System Design Evaluation , Hardware and Applications

COURSE OUTCOMES:

CO1: Visualize the objects in different dimensions.

CO2: Design and process the data for Visualization.

CO3: Apply the visualization techniques in physical sciences, computer science, applied mathematics and medical sciences.

CO4: Apply the virtualization techniques for research projects.

CO5: Identify appropriate data visualization techniques given particular requirements imposed by the data.

TOTAL: 45 PERIODS

REFERENCES

1. Matthew Ward, Georges Grinstein and Daniel Keim, “Interactive Data Visualization Foundations, Techniques, Applications”, 2010.
2. Colin Ware, “Information Visualization Perception for Design”, 4th edition, Morgan Kaufmann Publishers, 2021.
3. Robert Spence “Information visualization – Design for interaction”, Pearson Education, 2nd Edition, 2007.
4. Alexandru C. Telea, “Data Visualization: Principles and Practice,” A. K. Peters Ltd, 2008.

BD4001 PRINCIPLES OF SUPPLY CHAIN MANAGEMENT L T P C

3 0 0 3

COURSE OBJECTIVES:

- To analyze the supply chain scenario in India and its enabling factors.
- To make students to get insights on supply chain process from sourcing to distribution.
- To enhance the skills on supply chain integration and sustainable supply chain strategies among the students.
- To build knowledge in students to take care of any kinds of supply chain assignments in business organizations.

UNIT I INTRODUCTION 9

Role of Logistics and Supply chain Management: Scope and Importance- Evolution of Supply Chain - Decision Phases in Supply Chain - Competitive and Supply chain Strategies – Drivers of Supply Chain Performance and Obstacles.

UNIT II SUPPLY CHAIN NETWORK DESIGN 9

Role of Distribution in Supply Chain – Factors influencing Distribution network design – Design options for Distribution Network Distribution Network in Practice-Role of network Design in Supply Chain – Framework for network Decisions.

UNIT III LOGISTICS IN SUPPLY CHAIN 9

Role of transportation in supply chain – factors affecting transportations decision – Design option for transportation network – Tailored transportation – Routing and scheduling in transportation.

UNIT IV SOURCING AND COORDINATION IN SUPPLY CHAIN 9

Role of sourcing supply chain supplier selection assessment and contracts- Design collaboration - sourcing planning and analysis - supply chain coordination - Bullwhip effect – Effect of lack of coordination in supply chain and obstacles – Building strategic partnerships and trust within a supply chain.

UNIT V SUPPLY CHAIN AND INFORMATION TECHNOLOGY 9

The role IT in supply chain- The supply chain IT framework Customer Relationship Management – Internal supply chain management – supplier relationship management – future of IT in supply chain – E-Business in supply chain.

TOTAL : 45 PERIODS

COURSE OUTCOME:

Upon completion of the course, the student should be able to:

CO1: Understand the framework and scope of supply chain management.

CO2: Build and manage a competitive supply chain using strategies, models, techniques and information technology.

CO3: Analyze the logistics in supply chain

CO4:Plan the demand, inventory and supply and optimize supply chain Network

CO5: Evaluate the impact of IT on the Supply chain.

REFERENCES:

1. Sunil Chopra, Peter Meindl and Kalra, “Supply Chain Management, Strategy, Planning, and Operation”, Pearson Education,6th edition, 2015.
2. Jeremy F.Shapiro, “Modeling the Supply Chain”, Thomson Duxbury,2nd edition, 2006
3. David J.Bloomberg , Stephen Lemay and Joe B.Hanna, “Logistics”, PHI, 2002.
4. James B.Ayers, “Handbook of Supply Chain Management ", St.Lucie press,2nd edition, 2006.

BD4002 COMPUTATIONAL GEOMETRY L T P C

3 0 0 3

COURSE OBJECTIVES

- To understand geometric problems.
- To learn the algorithmic solutions for geometric problems.
- To learn the solutions for proximity problems
- To map problems in various application domains to a geometric problem.

UNIT I INTRODUCTION 9

Introduction – Application Domains – Line Segment Intersection – Intersection of Convex Polygons – Polygon Triangulation.

UNIT II GEOMETRIC SEARCHING 9

Geometric Searching – Range Searching – K- d-Trees – Range trees – Point-Location Problems.

UNIT III CONVEX HULL PROBLEM 9

Convex hull Problem – Preliminaries – Convex Hull Algorithms in the Plane – Graham‟s scan - Jarvis's March – Quick Hull – Divide-and-conquer – Dynamic Convex Hull Maintenance – Delaunay Triangulation.

UNIT IV PROXIMITY PROBLEMS 9

Proximity Problems – Fundamental Algorithms (Closest Pair – All Nearest Neighbours – Euclidean Minimum Spanning Tree – Nearest Neighbour Search) – Lower bounds – Closest Pair Problem : A Divide-and-Conquer Approach.

UNIT V VORONOI DIAGRAM 9

Voronoi Diagram – Proximity Problems Solved by the Voronoi Diagram – Planar Applications.

TOTAL : 45 PERIODS

COURSE OUTCOMES:

Upon completion of this course, the student should be able to

CO1: Transform problems in different applications to geometric problems

CO2: Use algorithms and techniques to solve search and point location problems

CO3: Understand and solve the complex hull problem

CO4: Solve proximity problems using various techniques

CO5: Use the appropriate and relevant, fundamental and applied computational knowledge, methodologies and modern tools in solving real -world problems.

REFERENCES:

1. Dr. Kalyanrao Takale , Dr. Shrikisan Gaikwad , Dr. Mrs. Nivedita Mahajan , Dr. Amjad Shaikh , Prof. Mrs. Shamal Deshmukh , Prof. S.R. Patil,1st Edition,,"Computational Geometry",2021.
2. David Mount,CMSC 754: Computational Geometry, 2021.Lecture notes from his Fall 2021 computational geometry course at Maryland.
3. Herbert Edelsbrunner, "Algorithms in Combinatorial Geometry, EATCS Monographs in Computer Science", Springer Verlag, 2011.

WEB REFRENCES:

1. *https://nptel.ac.in/courses/106/102/106102011/*
2. *https://ocw.mit.edu/courses/mechanical-engineering/2-158j-computational-geometry-spring-2003/*

ONLINE RESOURCES:

1. *https://www.hackerearth.com/practice/notes/computational-geometry-i-1/*
2. *https://algorithmtutor.com/Computational-Geometry/*

MP4093 **SOFT COMPUTING TECHNIQUES** **L T P C**
3 0 0 3

COURSE OBJECTIVES:

- To give the knowledge of soft computing theories fundamentals
- To provide the mathematical background for carrying out the optimization associated with neural network learning
- To familiarize the ideas of fuzzy sets, fuzzy logic, use of heuristics and Fuzzy Logic Control Systems
- To introduce the mathematical background for genetic algorithms
- To expose the hybrid soft computing systems and its applications

UNIT I SOFT COMPUTING FUNDAMENTALS 9

Introduction: Soft Computing Constituents – From Conventional Al to Computational Intelligence – Applications - Introduction, characteristics- learning methods - taxonomy - Evolution of neural networks - Artificial Neural Network (ANN): Fundamental Concept – Basic Terminologies – Neural Network Architecture – Learning Process – Fuzzy logic: Introduction – crisp - sets- fuzzy sets - crisp relations and fuzzy-relations: Cartesian product

UNIT II NEURAL NETWORKS 9

Fundamental Models of ANN: McCulloch- Pitts Model –Hebb Network – Linear Separability
Pitts Model –Hebb Network - Supervised Learning Networks: Perceptron Network – Adaline and Madaline Networks – Back Propagation Network – Radial Basis Function Network - Unsupervised Learning Networks: Kohonen Self Organizing Network – ART network - Hopfield Network - Special Network– Support Vector Machine- Kernel methods for Pattern classification- Kernel methods for function optimization.

UNIT III FUZZY COMPUTING AND MODELING 9

Fuzzy Equivalence and Tolerance Relation – Value assignments- Fuzzy Composition- Membership Functions–Fuzzification- Defuzzification: lambda cuts - Fuzzy Arithmetic – Extension Principle – Fuzzy Measures –Fuzzy Classification – Fuzzy Rules and Fuzzy Reasoning: Fuzzy Propositions – Formation of Rules – Decomposition of Rules – Aggregation of Rules – Approximate Reasoning – Fuzzy Inference and Expert Systems – Fuzzy Decision Making – Fuzzy Logic Control Systems.

UNIT IV GENETIC ALGORITHM AND APPLICATIONS 9

Genetic Algorithm: Fundamental Concept – Basic Terminologies – Traditional Vs Genetic Algorithm - Elements of GA - Encoding - Fitness Function – Genetic Operators: Reproduction – CrossOver - Inversion and Deletion - Mutation – Simple and General GA - The Schema Theorem- difference between GA and GP- Applications of GA. Multi-objective Optimization- Real-life case studies - optimization of traveling salesman problem using genetic algorithms

UNIT V HYBRID SOFT COMPUTING AND APPLICATIONS 9

Case Studies: Neuro-fuzzy Hybrid system- genetic neuro hybrid systems - genetic fuzzy hybrid and fuzzy genetic hybrid systems - simplified fuzzy ARTMAP – Applications: A fusion approach of multispectral images with SAR - Knowledge Leverage Based TSK Fuzzy System Modeling - Fuzzy C-Means algorithms for very large Data. Hybrid GA for Feature Selection- Multiobjective Genetic Fuzzy Clustering for pixel classification- Clustering Wireless Sensor Network Using Fuzzy Logic and Genetic Algorithm

COURSE OUTCOMES:

After completion of the course, the student will be able to:

CO1: Apply various soft computing concepts for practical applications

CO2: Choose and design suitable neural network for real time problems

CO3: Use fuzzy logic rules and reasoning to handle uncertainty and develop decision making and expert system

CO4: Describe the importance of genetic algorithms for solving combinatorial optimization problems

CO5: Analysis the various hybrid soft computing techniques and apply in real time problems

TOTAL: 45 PERIODS

REFERENCES:

1. S.N. Sivanandam, S.N. Deepa, "Principles of Soft Computing", Wiley, Second Edition, 2011.
2. S. Rajasekaran, G.A.V Vijayalakshmi Pai, "Neural Networks, Fuzzy Systems and Evolutionary Algorithms: Synthesis and Applications" Prentice Hall, Second Edition, 2017.
3. Timothy J. Ross, "Fuzzy Logic with Engineering Applications, 4th Edition, Wiley, 2016.
4. David E. Goldberg, Genetic Algorithm in Search Optimization and Machine Learning Pearson Education India, 2013.
5. Simon Haykin, Neural Networks Comprehensive Foundation Third Edition, Pearson Education, .2016.
6. James A. Freeman, David M. Skapura, Neural Networks Algorithms, Applications, and Programming Techniques, Pearson Education India, 2011.
7. J. -S. R. Jang, C.-T. Sun, E. Mizutani, "Neuro Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Pearson, 2015.

MU4152 MULTIMEDIA COMMUNICATION NETWORKS L T P C

3 0 0 3

COURSE OBJECTIVES:

- To recapitulate the fundamentals of networking and understand the requirements for multimedia communication.
- To learn guaranteed service model.
- To learn communication protocols that is frequently used in IoT ecosystems.
- To explore the support provided for multimedia communication in 3G and 4G networks.
- To study about VoIP and real time multimedia network applications.

UNIT I INTRODUCTION 9

Switched Networks and Shared media Networks – Circuit Switching, Packet Switching and Virtual Circuits – Flow Control and Congestion Control – TCP/IP reference model – Network Externalities – Service Integration – Elastic and Inelastic Traffic – Playback Applications – Additional Requirements For Inelastic Traffic – Core Networks And Access/Edge Networks.

Suggested Activities:

- Flipped classroom on network externalities and Economies of scale.
- External learning – Inter-continental backbone network and Autonomous Systems model of the Internet.
- Assignments on computing the playout time of packets.

Suggested Evaluation Methods:

- Quiz and discussion on network externalities and economies of scale.
- Assignments on proprietary protocols used in IoT and M2M.
- Assignments on problems related to playout time of multimedia applications.

UNIT II GUARANTEED SERVICE MODEL 9

Best Effort Service Model and Its Limitations – Qos Metrics – Diffserv and Intserv Networks – Queuing Techniques – WFQ and Its Variants – RED – Qos Aware Routing – Call Admission Control – RSVP – Policing and Traffic Shaping Algorithms – Multicast Routing – IGMP, Protocol Independent Multicast – PIM SM and PIM DM Variants.

Suggested Activities:

- Flipped classroom on IntServ and DiffServ networks.
- External learning – Exploring the ways of using DSCP in IP header.
- Assignments on finish time problems related to WFQ and its variants.

Suggested Evaluation Methods:

- Quiz and discussion on IntServ and DiffServ networks.
- Assignments on configuring a router in such a way that DSCP fielder is exploited to provide QoS.
- Assignments on problems related to the virtual finish and actual finish of packets in WFQ and its variants.

UNIT III MULTIMEDIA TRANSPORT 9

End To End Solutions – Laissez Faire Approach – Multimedia over TCP – Significance of UDP – Multimedia Streaming – Audio and Video Streaming – Accessing Audio And Video from a Web Server And Media Server – Removing Jitter at the Receiver – Recovering from Packet Loss – Forward Error Correction and Interleaving – Interactive And Non-Interactive Multimedia – Transcoding – RTSP – RTP/RTCP.

Suggested Activities:

- External learning – Exploring various media players available and the ways to customize them.
- Exploring the ways to configure RTP.
- Flipped classroom on RTP and RTCP.

Suggested Evaluation Methods:

- Assignments on media players available and configuring them.
- Configuring RTP and RTSP.
- Quiz and discussion on RTP and RTCP.

UNIT IV MULTIMEDIA OVER WIRELESS NETWORKS 9

Architecture of IP Multimedia Subsystem in 3G Networks – Application, Control and Data Planes in IMS Networks – Session Control, AAA, Real Time Data Transfer and Policy Control Protocols of IMS Networks – Relay Node and Multiple Radio Access Technologies in LTE – Voice Over IP Basics – IMS Volte Architecture – IP Multimedia Service Identity Module, Private Identity, Public Identity (ISIM, IMPI And IMPU) – SIP User Agent (SIP UAC And SIP UAE) – Real Time Polling Service and Extended Real Time Polling Service in IEEE 802.16/Wimax Networks.

Suggested Activities:

- Flipped classroom on IMSVoLTE architecture.
- External learning – Multimedia support in 5G networks.
- Analyzing the protocols of IP media subsystem.

Suggested Evaluation Methods:

- Quiz and discussion on IMSVoLTE architecture.
- Assignments on multimedia support in 5G networks.
- Assignments on analyzing the headers of IP multimedia subsystem.

UNIT V MULTIMEDIA NETWORKED APPLICATIONS 9

H.322 Standard – Protocol Stack And Call Setup – Session Initiation Protocol – Components, Messages And Operation – Supporting Protocols For SIP – Media Gateway Access Protocol, Resource Reservation Protocol, Session Description Protocol – Case Study – Video Conferencing – Military Surveillance – Interactive TV – Video On Demand – Smart Phone.

Suggested Activities:

- Flipped classroom on SCIBus and S.100.
- External learning – Multimedia access networks and edge networks.
- Exploring the ways to configure SIP.

Suggested Evaluation Methods:

- Quiz and discussion on SCIBus and S.100.
- Assignments on multimedia access networks and edge networks.
- Configuring SIP using suitable commands.

TOTAL: 45 PERIODS

COURSE OUTCOMES:

On completion of the course, the students will be able to:

CO1: Deploy the right multimedia communication models.

CO2: Apply QoS to multimedia network applications at the network level with efficient scheduling and routing techniques.

CO3: Apply QoS to multimedia network applications at the end system level with efficient scheduling and routing techniques.

CO4: Understand IP multimedia subsystem and IP initiatives in cellular networks to support multimedia traffic.

CO5: Design and implement VoIP based solutions for multimedia transport.

CO6: Develop the real-time multimedia network applications.

REFERENCES:

1. Mario Marques da Silva, “Multimedia Communications and Networking”, CRC Press, 2012
2. K. R. Rao, Zoron S. Bojkovic, Bojan M. Bakmaz, “Wireless Multimedia Communication Systems: Design, Analysis and Implementation”, CRC Press, 2017
3. Jim Kurose, Keith Ross, “Computer Networking: A Top Down Approach”, Pearson Education, 2017
4. K. R. Rao, Zoran S. Bojkovic, Dragorad A. Milovanovic, “Introduction to Multimedia Communications Applications, Middleware, Networking”, John Wiley and Sons, 2009

BD4003 PARALLEL AND DISTRIBUTED COMPUTING L T P C
3 0 0 3

COURSE OBJECTIVES:

- To select the appropriate parallel programming model for the given application.
- To practice the constructs of a parallel programming model to convert a sequential program to parallel program.
- To develop parallel programs using Open MP and MPI constructs.
- To study software components of distributed computing systems. Know about the communication and interconnection architecture of multiple computer systems.
- To recognize the inherent difficulties that arise due to distributedness of computing resources.
- To understand the concepts of networks & protocols, mobile & wireless computing and their applications in solving real world problems.

UNIT I FUNDAMENTALS 9

Introduction to Parallel and Distributed Computing - Parallel and Distributed Architectures, Socket programming - Parallel Performance, Shared Memory and Threads – Parallel Algorithms - Parallel Algorithms, OpenMP - Scalable Algorithms, Message Passing - Parallel,work sharing and Synchronization constructs. MPI data types and tags, environment management routines. Data level parallelism - Vector architecture- Graphics processing units-Programming the GPU.

UNIT II PROCESS AND DISTRIBUTED OBJECTS 9

Characterization of Distributed Systems – Examples – Resource Sharing and the Web – Challenges – System Models – Architectural and Fundamental Models – Networking and Internetworking – Types of Networks – Network Principles – External Data Representation and Marshalling – Client-–Server Communication – Group Communication – Case Study –Distributed Objects and Remote Invocation – Communication between Distributed Objects –Remote Procedure Call.

UNIT III OPERATING SYSTEM ISSUES I 9

The OS Layer – Protection – Processes and Threads – Communication and Invocation – OS Architecture – Security – Overview – Cryptographic Algorithms – Digital Signatures –Cryptography Pragmatics – Case Studies – Distributed File Systems – File Service Architecture.

UNIT IV OPERATING SYSTEM ISSUES II 9

Name Services – Domain Name System – Directory and Discovery Services – Global Name Service – X.500 Directory Service – Clocks – Events and Process States – Synchronizing Physical Clocks – Logical Time and Logical Clocks – Global States – Distributed Debugging– Distributed Mutual Exclusion – Elections – Multicast Communication Related Problems.

UNIT V DISTRIBUTED TRANSACTION PROCESSING 9

Transactions – Nested Transactions – Locks – Optimistic Concurrency Control – Timestamp Ordering – Comparison – Flat and Nested Distributed Transactions – Atomic Commit Protocols – Concurrency Control in Distributed Transactions – Distributed Deadlocks –Transaction Recovery – Overview of Replication and Distributed Multimedia Systems.

TOTAL: 45 PERIODS

COURSE OUTCOMES:

Upon completion of the course, the student should be able to:

CO1: Work on appropriate parallel programming model

CO2: Model and visualize Design protocols for ensuring cache coherence
CO3: Students will be familiar with the design of the Distributed system.
CO4: To implementation and security issues of distributed systems.
CO5: To implement Transaction Processing of distributed systems.

REFERENCES:

1. David E. Culler, Jaswinder Pal Singh, "Parallel Computing Architecture : A Hardware/Software Approach ", Morgan Kaufmann / Elsevier Science (reprint Technical Science & Engineering) , 2nd Edition,1998.
2. George Coulouris, Jean Dollimore and Tim Kindberg, "Distributed Systems Concepts and Design", Pearson Education, 5th Edition, 2017.
3. SapeMullender, "Distributed Systems", Addison Wesley, 2nd Edition, 1993.
4. Brendan Burns, " Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable Services",O′Reilly, 1st edition,2018.

WEB REFERENCES:

1. *https://moodle.risc.jku.at/course/view.php?id=143*

ONLINE RESOURCES:

1. *https://www.youtube.com/watch?v=dlBVWMdGhqw&list=PLUJ7JmcrTifBROWODSG8wgyl20XgBuE-N*

2. *https://www.youtube.com/watch?v=qbQCQ0U6H0o&list=PLbMVogVj5nJQRvzENlvMKA9q70ScSRZBQ*

BD4091 PREDICTIVE MODELLING L T P C
3 0 0 3

COURSE OBJECTIVES:

- To understand the terms and terminologies of predictive modeling.
- To study the various predictive models, their merits, demerits and application.
- To get exposure to various analytical tools available for predictive modeling.
- To learn the predictive modeling markup language.
- To get familiar with the technologies in predictive modeling.

UNIT I INTRODUCTION TO PREDICTIVE MODELING 9

Core ideas in data mining - Supervised and unsupervised learning - Classification vs. Prediction - Steps in data mining- SEMMA Approach - Sampling -Pre-processing - Data cleaning - Data Partitioning - Building a model - Statistical models - Statistical models for predictive analytics.

UNIT II PREDICTIVE MODELING BASICS 9

Data splitting – Balancing- Over fitting –Oversampling –Multiple Regression Artificial neural networks (MLP) - Variable importance- Profit/loss/prior probabilities - Model specification - Model selection - Multivariate Analysis.

UNIT III PREDICTIVE MODELS 9

Association Rules-Clustering Models –Decision Trees- Ruleset Models- KNearest Neighbors – Naive Bayes - Neural Network Model – Regression Models – Regression Trees – Classification & Regression Trees (CART) – Logistic Regression – Multiple Linear Regression Scorecards – Support Vector Machines – Time Series Models - Comparison between models - Lift chart Assessment of a single model.

UNIT IV PREDICTIVE MODELING MARKUP LANGUAGE 9

Introduction to PMML – PMML Converter - PMML Structure – Data Manipulation in PMML – PMML Modeling Techniques - Multiple Model Support – Model Verification.

UNIT V TECHNOLOGIES AND CASE STUDIES 9

Weka – RapidMiner – IBM SPSS Statistics- IBM SPSS Modeler – SAS Enterprise Miner – Apache Mahout – R Programming Language.-Real time case study with modeling and analysis.

TOTAL: 45 PERIODS

COURSE OUTCOMES:

Upon completion of the course, the student should be able to:

CO1: Design and analyze appropriate predictive models.
CO2: Define the predictive models using PMML.
CO3: Apply statistical tools for analysis.
CO4: Use various analytical tools available for predictive modeling.
CO5: Apply predictive modeling markup language in data manipulation .

REFERENCES:

1. Kattamuri S. Sarma, “Predictive Modeling with SAS Enterprise Miner: Practical Solutions for Business Applications”, 3rd Edition, SAS Publishing, 2017.
2. Alex Guazzelli, Wen-Ching Lin, Tridivesh Jena, James Taylor, “PMML in Action Unleashing the Power of Open Standards for Data Mining and Predictive Analytics”, 2nd Edition, Create Space Independent Publishing Platform,2012.
3. Ian H. Witten, Eibe Frank , “Data Mining: Practical Machine Learning Tools and Techniques”, Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 3rd Edition, 2011.
4. Eric Siegel , “Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die”, 2nd Edition, Wiley, 2016.
5. Conrad Carlberg, “Predictive Analytics: Microsoft Excel”, 1st Edition, Que Publishing, 2012.
6. Jeremy Howard, Margit Zwemer, Mike Loukides, “Designing Great Data Products- Inside the Drivetrain train Approach, a Four-Step Process for Building Data Products – Ebook”, 1st Edition, O'Reilly Media, March 2012.

WEB REFERENCES:

1. *https://nptel.ac.in/courses/108108111/*
2. *https://www.coursera.org/learn/predictive-modeling-analytics*

ONLINE RESOURCES:

1. *https://bookdown.org/egarpor/PM-UC3M/*
2. *https://cics.nd.edu/research/applications/materials/*

BD4004 IMAGE PROCESSING AND ANALYSIS L T P C
3 0 0 3

COURSE OBJECTIVES:

- To understand the basics of digital images and noise models
- To understand spatial domain filters and frequency domain filters
- To understand the image processing techniques
- To familiarize the image processing environment and their applications
- To appreciate the use of image processing in various applications

UNIT I SPATIAL DOMAIN PROCESSING 9

Introduction to image processing – imaging modalities – image file formats – image sensing and acquisition – image sampling and quantization – noise models – spatial filtering operations – histograms – smoothing filters – sharpening filters – fuzzy techniques for spatial filtering – spatial filters for noise removal.

UNIT II FREQUENCY DOMAIN PROCESSING 9

Frequency domain – Review of Fourier Transform (FT), Discrete Fourier Transform (DFT), and Fast Fourier Transform (FFT) – filtering in frequency domain – image smoothing – image sharpening – selective filtering – frequency domain noise filters wavelets – Haar Transform – multiresolution expansions – wavelet transforms wavelets based image processing.

UNIT III SEGMENTATION AND EDGE DETECTION 9

Thresholding techniques – region growing methods – region splitting and merging adaptive thresholding – threshold selection – global valley – histogram concavity edge detection –template matching – gradient operators – circular operators differential edge operators –hysteresis thresholding – Canny operator – Laplacian operator – active contours – object segmentation.

UNIT IV INTEREST POINTS, MORPHOLOGY, AND TEXTURE 9

Corner and interest point detection – template matching – second order derivatives median filter based detection – Harris interest point operator – corner orientation local invariant feature detectors and descriptors – morphology – dilation and erosion morphological operators – grayscale morphology – noise and morphology – texture texture analysis – co-occurrence matrices – Laws' texture energy approach – Ade's eigen filter approach.

UNIT V COLOR IMAGES AND IMAGE COMPRESSION 9

Color models – pseudo colors – full-color image processing – color transformations smoothing and sharpening of color images – image segmentation based on color noise in color images. Image Compression – redundancy in images – coding redundancy – irrelevant information in images – image compression models – basic compression methods – digital image watermarking.

TOTAL : 45 PERIODS

COURSE OUTCOMES:

CO1: Design and implement algorithms for image processing applications that incorporates different concepts of medical Image Processing

CO2: Explain image modalities, sensing, acquisition, sampling, and quantization, noise models and implement spatial filter operations

CO3: Familiar with the use of MATLAB and its equivalent open source tools

CO4: Critically analyze different approaches to image processing applications

CO5: Explore the possibility of applying Image processing concepts in various applications

REFERENCES:

1. E. R. Davies, “Computer & Machine Vision”, Fourth Edition, Academic Press, 2012.
2. S. Sridhar, “Digital Image Processing”, 2nd Edition, Oxford University Press, 2016.
3. W. Burger and M. Burge, “Digital Image Processing: An Algorithmic Introduction using Java”, Springer,2nd edition, 2016.
4. John C. Russ, “The Image Processing Handbook”, Sixth Edition, CRC Press, 2011.
5. R. C. Gonzalez and R. E. Woods, “Digital Image Processing”, Third Edition, Pearson,2008.
6. Mark Nixon and Alberto S. Aquado, “Feature Extraction & Image Processing for Computer Vision”, Third Edition, Academic Press, 2013.

7. D. L. Baggio et al., “Mastering OpenCV with Practical Computer Vision Projects”, Packt Publishing, 2012.
8. Jan Erik Solem, “Programming Computer Vision with Python: Tools and algorithms for analyzing images”, O'Reilly Media, 2012.

MP4091 **COGNITIVE COMPUTING** **L T P C**
3 0 0 3

COURSE OBJECTIVES:

- To familiarize Use the Innovation Canvas to justify potentially successful products.
- To learn various ways in which to develop a product idea.
- To understand about how Big Data can play vital role in Cognitive Computing
- To know about the business applications of Cognitive Computing
- To get into all applications of Cognitive Computing

UNIT I FOUNDATION OF COGNITIVE COMPUTING 9

Foundation of Cognitive Computing: cognitive computing as a new generation, the uses of cognitive systems, system cognitive, gaining insights from data, Artificial Intelligence as the foundation of cognitive computing, understanding cognition Design Principles for Cognitive Systems: Components of a cognitive system, building the corpus, bringing data into cognitive system, machine learning, hypotheses generation and scoring, presentation, and visualization services

UNIT II NATURAL LANGUAGE PROCESSING IN COGNITIVE SYSTEMS 9

Natural Language Processing in support of a Cognitive System: Role of NLP in a cognitive system, semantic web, Applying Natural language technologies to Business problems Representing knowledge in Taxonomies and Ontologies: Representing knowledge, Defining Taxonomies and Ontologies, knowledge representation, models for knowledge representation, implementation considerations

UNIT III BIG DATA AND COGNITIVE COMPUTING 9

Relationship between Big Data and Cognitive Computing: Dealing with human-generated data, defining big data, architectural foundation, analytical data warehouses, Hadoop, data in motion and streaming data, integration of big data with traditional data Applying Advanced Analytics to cognitive computing: Advanced analytics is on a path to cognitive computing, Key capabilities in advanced analytics, using advanced analytics to create value, Impact of open source tools on advanced analytics

UNIT IV BUSINESS IMPLICATIONS OF COGNITIVE COMPUTING 9

Preparing for change ,advantages of new disruptive models , knowledge meaning to business, difference with a cognitive systems approach , meshing data together differently, using business knowledge to plan for the future , answering business questions in new ways , building business specific solutions , making cognitive computing a reality , cognitive application changing the market The process of building a cognitive application: Emerging cognitive platform, defining the objective, defining the domain, understanding the intended users and their attributes, questions and exploring insights, training and testing

UNIT V APPLICATION OF COGNITIVE COMPUTING 9

Building a cognitive health care application: Foundations of cognitive computing for healthcare, constituents in healthcare ecosystem, learning from patterns in healthcare Data, Building on a foundation of big data analytics, cognitive applications across the health care eco system, starting with a cognitive application for healthcare, using cognitive applications to improve health and wellness, using a cognitive application to enhance the electronic medical record Using cognitive application to improve clinical teaching

COURSE OUTCOMES:

CO1: Explain applications in Cognitive Computing.
CO2: Describe Natural language processor role in Cognitive computing.
CO3: Explain future directions of Cognitive Computing
CO4: Evaluate the process of taking a product to market
CO5: Comprehend the applications involved in this domain.

TOTAL:45 PERIODS

REFERENCES

1. Judith H Hurwitz, Marcia Kaufman, Adrian Bowles, “Cognitive computing and Big Data Analytics”, Wiley, 2015
2. Robert A. Wilson, Frank C. Keil, “The MIT Encyclopedia of the Cognitive Sciences”, The MIT Press, 1999.
3. Noah D. Goodman, Joshua B. Tenenbaum, The ProbMods Contributors, “Probabilistic Models of Cognition”, Second Edition, 2016, https://probmods.org/.

IF4095 **SOCIAL NETWORK ANALYSIS** **L T P C**
3 0 0 3

COURSE OBJECTIVES:

- Formalise different types of entities and relationships as nodes and edges and represent this information as relational data.
- Understand the fundamental concepts in analyzing the large-scale data that are derived from social networks
- Understand the basic concepts and principles of different theoretical models of social networks analysis.
- Transform data for analysis using graph-based and statistics-based social network measures
- Choose among social network designs based on research goals

UNIT I GRAPH THEORY AND STRUCTURE 10

Breadth First Search (BFS) Algorithm. Strongly Connected Components (SCC) Algorithm. Weakly Connected Components (WCC) Algorithm. First Set of Experiments—Degree Distributions. Second Set of Experiments—Connected Components. Third Set of Experiments—Number of Breadth First Searches. Rank Exponent R. Out-Degree Exponent O. Hop Plot Exponent H. Eigen Exponent E. Permutation Model. Random Graphs with Prescribed Degree Sequences. Switching Algorithms. Matching Algorithm. “Go with the Winners” Algorithm. HyperANF Algorithm. Iterative Fringe Upper Bound (iFUB) Algorithm. Spid. Degree Distribution. Path Length. Component Size. Clustering Coefficient and Degeneracy. Friends-of-Friends. Degree Assortativity. Login Correlation.

UNIT II SOCIAL NETWORK GRAPH ANALYSIS 9

Social network exploration/ processing and properties: Finding overlapping communities, similarity between graph nodes, counting triangles in graphs, neighborhood properties of graphs. Pregel paradigm and Apache Giraph graph processing system.

UNIT III INFORMATION DIFFUSION IN SOCIAL NETWORKS 9

Strategic network formation: game theoretic models for network creation/ user behavior in social networks. Information diffusion in graphs: Cascading behavior, spreading, epidemics, heterogeneous social network mining, influence maximization, outbreak detection. Opinion analysis on social networks: Contagion, opinion formation, coordination and cooperation.

UNIT IV CASCADING IN SOCIAL NETWORKS 8

Cascading in Social Networks. Decision Based Models of Cascade. Collective Action. Cascade Capacity. Co-existence of Behaviours. Cascade Capacity with Bilinguality. Probabilistic Models of Cascade. Branching Process. Basic Reproductive Number. SIR Epidemic Model. SIS Epidemic Model. SIRS Epidemic Model. Transient Contact Network. Cascading in Twitter.

UNIT V LINK ANALYSIS & COMMUNITY DETECTION 9

Search Engine. Crawling. Storage. Indexing. Ranking. Google. Data Structures. Crawling. Searching. Web Spam Pages Strength of Weak Ties. Triadic Closure. Detecting Communities in a Network. Girvan-Newman Algorithm. Modularity. Minimum Cut Trees. Tie Strengths in Mobile Communication Network. Exact Betweenness Centrality. Approximate Betweenness Centrality.

SUGGESTED ACTIVITIES:

- Twitter Intelligence project performs tracking and analysis of the Twitter
- Large-Scale Network Embedding as Sparse Matrix Factorization
- Implement how Information Propagation on Twitter
- Social Network Analysis and Visualization software application.
- Implement the Structure of Links in Networks

COURSE OUTCOMES:

CO1: Plan and execute network analytical computations.
CO2: Implement mining algorithms for social networks
CO3: Analyze and evaluate social communities.
CO4: Use social network analysis in behavior analytics
CO5: Perform mining on large social networks and illustrate the results.

TOTAL : 45 PERIODS

REFERENCES

1. Practical Social Network Analysis with Python, Krishna Raj P. M. Ankith Mohan and K. G. Srinivasa. Springer, 2018
2. SOCIAL NETWORK ANALYSIS: METHODS AND APPLICATIONS, STANLEY WASSERMAN, and KATHERINE F' AUST. CAMBRIDGE UNIVERSITY PRESS, 2012
3. Social Network Analysis: History, Theory and Methodology by Christina Prell, SAGE Publications, 1st edition, 2011
4. Sentiment Analysis in Social Networks, Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, and Bing. LiuElsevier Inc, 1st edition, 2016
5. Social Network Analysis, John Scott. SAGE Publications, 2012

BD4005 VIRTUALIZATION TECHNIQUES AND APPLICATIONS L T P C

3 0 0 3

COURSE OBJECTIVES:

- To understand and deploy practical virtualization solutions and enterprise solutions.
- Able to understand Server Virtualization and Virtualization Platform.
- Understand the technologies of Virtualization and Network Virtualization.
- To understand the concepts of Virtualization storage.
- To study virtual machine products.

UNIT I OVERVIEW OF VIRTUALIZATION 9

System architectures - Virtual Machine basics - Process vs System Virtual Machines - Taxonomy. Emulation: Basic Interpretation - Threaded Interpretation - Precoded and Direct Threaded Interpretation - Binary Translation. System Virtual Machines - Key concepts - Resource utilization basics.

UNIT II PROCESS VIRTUAL MACHINES 9

Implementation – Compatibility – Levels – Framework – State Mapping – Register – Memory Address Space – Memory Architecture Emulation – Memory Protection – Instruction Emulation – Performance Tradeoff - Staged Emulation – Exception Emulation – Exception Detection – Interrupt Handling – Operating Systems Emulation – Same OS Emulation – Different OS Emulation – System Environment .

UNIT III HIGH LEVEL LANGUAGE VIRTUAL MACHINES AND SERVER VIRTUALIZATION 9

HLL virtual machines: Pascal P-Code – Object Oriented HLLVMs - Java VM architecture - Java Native Interface - Common Language Infrastructure. Server virtualization: Partitioning techniques - virtual hardware - uses of virtual servers - server virtualization platforms.

UNIT IV NETWORK AND STORAGE VIRTUALIZATION 9

Design of Scalable Enterprise Networks – Layer2 Virtualization – VLAN - VFI - Layer 3 Virtualization – VRF - Virtual Firewall Contexts - Network Device Virtualization - Data- Path Virtualization - Routing Protocols. Hardware Devices – SAN backup and recovery techniques – RAID – Classical Storage Model – SNIA Shared Storage Model – Virtual Storage: File System Level and Block Level.

UNIT V APPLYING VIRTUALIZATION 9

Practical Virtualization Solutions: Comparison of Virtualization Technologies: Guest OS/ Host OS – Hypervisor – Emulation – Kernel Level – Shared Kernel, Enterprise Solutions: VMWare Server – VMWare ESXi – Citrix Xen Server – Microsoft Virtual PC – Microsoft Hyper-V – Virtual Box, Server Virtualization: Configuring Servers with Virtualization – Adjusting and Tuning Virtual servers – VM Backup – VM Migration, Desktop Virtualization: Terminal services – Hosted Desktop – Web-based Solutions – Localized Virtual Desktops, Network and Storage Virtualization: Virtual Private Networks – Virtual LAN – SAN and VSAN – NAS.

TOTAL : 45 PERIODS

COURSE OUTCOMES:

Upon completion of this course, the student should be able to

CO1: Design and develop applications on virtual machine platforms

CO2:To understand the concepts of virtualization and virtual machines.

CO3:To understand the implementation of process and system virtual machines.

CO4: To explore the aspects of high level language virtual machines.
CO5:To gain expertise in server, network and storage virtualization.

REFERENCES:

1. William von Hagen, "Professional Xen Virtualization", Wrox Publications, January, 2008.
2. James E. Smith, Ravi Nair, “Virtual Machines: Versatile Platforms for Systems and Processes”, Elsevier/Morgan Kaufmann, 2012
3. David Marshall, Wade A. Reynolds, “Advanced Server Virtualization: VMware and Microsoft Platform in the Virtual Data Center”, Auerbach Publications, 2006.
4. Kumar Reddy, Victor Moreno, “Network virtualization”, Cisco Press, July, 2006.
5. Chris Wolf, Erick M. Halter, “Virtualization: From the Desktop to the Enterprise”, APress 2005.
6. Kenneth Hess , Amy Newman, “Practical Virtualization Solutions: Virtualization from the Trenches”, Prentice Hall, 2010

WEB REFERENCES:

1. lecturer.eepis-its.edu/ ~isbat/materikuliah/ .../ Virtualization. ppt
2. cs.nju.edu.cn/distribute-systems/lecture-notes/c11.ppt
3. https://edux.fit.cvut.cz/oppa/MI-POA/prednasky/MI-POA10.pdf
4.www.cs.otago.ac.nz/cosc440/lectures/lecture%2010.ppt

ONLINE RESOURCES:

1. https://www.simplilearn.com/virtualization
2. tutorialspoint.com/virtualization2.0/virtualization2.0_overview.htm

ML4291 NATURAL LANGUAGE PROCESSING L T P C
2 0 2 3

COURSE OBJECTIVES:

- To understand basics of linguistics, probability and statistics
- To study statistical approaches to NLP and understand sequence labeling
- To outline different parsing techniques associated with NLP
- To explore semantics of words and semantic role labeling of sentences
- To understand discourse analysis, question answering and chatbots

UNIT I INTRODUCTION 6

Natural Language Processing – Components - Basics of Linguistics and Probability and Statistics – Words-Tokenization-Morphology-Finite State Automata

UNIT II STATISTICAL NLP AND SEQUENCE LABELING 6

N-grams and Language models –Smoothing -Text classification- Naïve Bayes classifier – Evaluation - Vector Semantics – TF-IDF - Word2Vec- Evaluating Vector Models -Sequence Labeling **–** Part of Speech – Part of Speech Tagging -Named Entities **–**Named Entity Tagging

UNIT III CONTEXTUAL EMBEDDING 6

Constituency –Context Free Grammar –Lexicalized Grammars- CKY Parsing – Earley's algorithm- Evaluating Parsers -Partial Parsing – Dependency Relations- Dependency Parsing -Transition Based - Graph Based

UNIT IV COMPUTATIONAL SEMANTICS 6

Word Senses and WordNet – Word Sense Disambiguation – Semantic Role Labeling – Proposition Bank- FrameNet- Selectional Restrictions - Information Extraction - Template Filling

UNIT V DISCOURSE ANALYSIS AND SPEECH PROCESSING 6

Discourse Coherence – Discourse Structure Parsing – Centering and Entity Based Coherence – Question Answering –Factoid Question Answering – Classical QA Models – Chatbots and Dialogue systems – Frame-based Dialogue Systems – Dialogue–State Architecture

30 PERIODS

SUGGESTED ACTIVITIES:

1. **1.** Probability and Statistics for NLP Problems
2. Carry out Morphological Tagging and Part-of-Speech Tagging for a sample text
3. Design a Finite State Automata for more Grammatical Categories
4. Problems associated with Vector Space Model
5. Hand Simulate the working of a HMM model
6. Examples for different types of work sense disambiguation
7. **7.** Give the design of a Chatbot

PRACTICAL EXERCISES: 30 PERIODS

1. Download nltk and packages. Use it to print the tokens in a document and the sentences from it.
2. Include custom stop words and remove them and all stop words from a given document using nltk or spaCY package
3. Implement a stemmer and a lemmatizer program.
4. Implement asimple Part-of-Speech Tagger
5. Write a program to calculate TFIDF of documents and find the cosine similarity between any two documents.
6. Use nltk to implement a dependency parser.
7. Implement a semantic language processor that uses WordNet for semantic tagging.
8. Project - (in Pairs) Your project must use NLP concepts and apply them to some data.
    a. Your project may be a comparison of several existing systems, or it may propose a new system in which case you still must compare it to at least one other approach.
    b. You are free to use any third-party ideas or code that you wish as long as it is publicly available.
    c. You must properly provide references to any work that is not your own in the write-up.
    d. Project proposal You must turn in a brief project proposal.
    e. Your project proposal should describe the idea behind your project. You should also briefly describe software you will need to write, and papers (2-3) you plan to read.

List of Possible Projects

1. Sentiment Analysis of Product Reviews
2. Information extraction from News articles
3. Customer support bot
4. Language identifier
5. Media Monitor
6. Paraphrase Detector
7. Identification of Toxic Comment
8. Spam Mail Identification

COURSE OUTCOMES:

CO1: Understand basics of linguistics, probability and statistics associated with NLP

CO 2: Implement a Part-of-Speech Tagger

CO 3: Design and implement a sequence labeling problem for a given domain

CO 4: Implement semantic processing tasks and simple document indexing and searching system using the concepts of NLP

CO 5: Implement a simple chatbot using dialogue system concepts

TOTAL : 60 PERIODS

REFERENCES

1. Daniel Jurafsky and James H.Martin, “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition” (Prentice Hall Series in Artificial Intelligence), 2020
2. Jacob Eisenstein. “Natural Language Processing “, MIT Press, 2019
3. Samuel Burns “Natural Language Processing: A Quick Introduction to NLP with Python and NLTK, 2019
4. Christopher Manning, “Foundations of Statistical Natural Language Processing”, MIT Press, 2009.
5. Nitin Indurkhya,Fred J. Damerau, “Handbook of Natural Language Processing”, Second edition, Chapman & Hall/CRC: Machine Learning & Pattern Recognition, Hardcover,2010
6. Deepti Chopra, Nisheeth Joshi, “Mastering Natural Language Processing with Python”, Packt Publishing Limited, 2016
7. Mohamed Zakaria Kurdi “Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax (Cognitive Science)”, ISTE Ltd., 2016
8. Atefeh Farzindar,Diana Inkpen, “Natural Language Processing for Social Media (Synthesis Lectures on Human Language Technologies)”, Morgan and Claypool Life Sciences, 2015

BD4006 DATA INTENSIVE COMPUTING

L	T	P	C
3	0	0	3

COURSE OBJECTIVES:

- To understand the basics of the various database systems including databases for Big data.
- To learn the architecture of data intensive computing.
- To learn parallel processing for data intensive computing.
- To learn Security in Data Intensive Computing Systems.
- To learn the applications that involve Data intensive computing.

UNIT I INTRODUCTION 9

Introduction to Distributed systems – Databases Vs. File Systems - Distributed file systems(HDFS) – Distributed Machine-Learning System - Data Parallelism – Characteristics -Hadoop –Execution Engines -Map Reduce- Distributed Storage System for Structured Data – NoSQL databases - Casandra, Mongo DB-Developing a Distributed Application

UNIT II ARCHITECTURES AND SYSTEMS 9

High performance Network Architectures for Data intensive Computing – Architecting Data Intensive Software systems – ECL/HPCC: A Unified approach to Big Data – Scalable storage for Data Intensive Computing - Computation and Storage of scientific data sets in cloud- Stream Data Model - Architecture for Data Stream Management-Stream Queries –Sampling Data in a Stream Filtering Streams

UNIT III TECHNOLOGIES AND TECHNIQUES 9

Load balancing techniques for Data Intensive computing – Resource Management for Data Intensive Clouds – SALT - Parallel Processing, Multiprocessors and Virtualization in Data intensive Computing - Challenges in Data Intensive Analysis and Visualization - Large-Scale Data Analytics Using Ensemble Clustering - Ensemble Feature Ranking Methods for Data Intensive Computing Application - Record Linkage Methodology and Applications- Semantic Wrapper

UNIT IV SECURITY 9

Security in Data Intensive Computing Systems - Data Security and Privacy in Data-Intensive Supercomputing Clusters - Information Security in Large Scale Distributed Systems -Privacy and Security Requirements of Data Intensive Applications in Clouds

UNIT V APPLICATIONS AND FUTURE TRENDS 9

Cloud and Grid Computing for Data Intensive Applications -Scientific Applications - Bioinformatics Large Science Discoveries - Climate Change - Environment - Energy - Commercial Applications - Future trends in Data Intensive Computing

TOTAL : 45 PERIODS

COURSE OUTCOMES:

Upon completion of the course, the students will be able to

CO1: Design applications that involve data intensive computing.

CO2: Suggest appropriate architecture for data intensive computing systems.

CO3: Decide on the appropriate techniques of Map Reduce, Mongo DB, for the different Applications.

CO4: Identify parallel processing techniques for data intensive computing.

CO5: Decide on the various security techniques that are necessary for data intensive applications.

REFERENCES:

1. Tom White, “Hadoop: The Definitive Guide”, O'Reilly Media, 4th edition,2015.
2. Hector Garcia-Molina, Jeffrey D. Ullman, and Jennifer Widom., “Database Systems: The Complete Book”, Pearson, 2013.
3. Furht, Borko, Escalante, Armando, “Handbook of Data Intensive Computing”, Springer 2011.

WEB REFERENCES:

1. *https://en.wikipedia.org/wiki/Data-intensive_computing*

2. https://www.computer.org/csdl/magazine/co/2008/04/mco2008040030/13rRUNvgyZ8

ONLINE RESOURCES:

1. https://www.slideshare.net/huda2018/dataintensive-technologies-for-cloudcomputing

BD4007 R LANGUAGE FOR MINING L T P C
3 0 0 3

COURSE OBJECTIVES

- To study the major data mining problems as different types of computational tasks (prediction, classification, clustering, etc.) and the algorithms appropriate for addressing these tasks
- To learn how to analyze data through statistical and graphical summarization, supervised and unsupervised learning algorithms
- To systematically evaluate data mining algorithms and understand how to choose algorithms for different analysis tasks

UNIT I INTRODUCTION DATA MINING 9

Introduction, Mining Association Rules in Large Databases, Mining Frequent Patterns - basic concepts - Efficient and scalable frequent item set mining methods, Apriori algorithm, FP-Growth algorithm, Associations - mining various kinds of association rules.

UNIT II PREDICTIVE MODELING AND CLUSTERING 9

Classification and Prediction-Issues Classification by Decision Tree Induction–Bayesian Classification – Other Classification Methods – Prediction–Clusters Analysis – Basics of cluster analysis -Types of Data in Cluster Analysis – Categorization of Major Clustering Methods – Partitioning Methods – Hierarchical Methods.

UNIT III MINING DATA STREAMS 9

Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time Analytics Platform(RTAP) Applications.

UNIT IV DATA ANALYTIC METHODS USING R 9

Introduction to R – Analyzing and exploring data with R - Statistics for model building and evaluation. Analytical Approaches, introducing to various Analytical Tools, Installing R, Handling Basic Expressions in R, Variables in R, Working with Vectors, Storing and Calculating Values in R, Creating and Using Objects, Interacting with Users, Handling Data in R Workspace

UNIT V FUNCTIONS AND PACKAGES IN R 9

Executing Scripts, Reading Datasets and Exporting Data from Manipulating and Processing Data in R, Working with Functions and Packages in R, Performing Graphical Analysis in R, Techniques Used for Visual Data Representation, Types of Data Visualization.

TOTAL: 45 PERIODS

COURSE OUTCOMES:

Upon completion of the course, the student should be able to:

CO1: Demonstrate accurate and efficient use of classification using the R system for the computations.

CO2: Demonstrate the related data mining techniques Using R
CO3: Demonstrate capacity for mathematical reasoning through analyzing, proving and explaining concepts from the theory that underpins classification and related data mining methods
CO4: Apply problem-solving using classification and related data mining techniques to diverse situations in business, biology, engineering and other sciences
CO5: Analyze the data visualization

REFERENCES:

1. Carlo Vercellis, Business Intelligence: Data mining and Optimization for Decision Making, WILEY.
2. Han J., Kamber M. and Pei J, Data mining concepts and techniques, Morgan Kaufmann Publishers (2011) 3rd ed.
3. Big Data Computing and Communications edited by Yu Wang, Hui Xiong, ShlomoArgamon, XiangYang Li, JianZhong Li Springer
4. Andrea Cirillo,"R Data Mining: Implement data mining techniques through practical use cases and real world datasets",Packt Publication,1st Edition,2017.
5. Luis Torgo."Data Mining with R" Learning with Case Studies, Second Edition 2020,Chapman and Hall/CRC.

WEB REFERENCES:

1. *https://onlinecourses-archive.nptel.ac.in/noc18-mg11/announcements*
2. *https://swayam.gov.in/nd1_noc19_ma33/preview*
3. *www.datacamp.com/**R**-Tutorial*

ONLINE RESOURCES:

1. *https://www.youtube.com/watch?v=BB2O4VCu5j8*
2. *https://www.tutorialspoint.com/r/index.htm*
3. *http://www.rdatamining.com/*

CP4097 WEB ANALYTICS L T P C
3 0 0 3

COURSE OBJECTIVES:

- To understand the Web analytics platform, and their evolution.
- To learn about the various Data Streams Data.
- To learn about the benefits of surveys and capturing of data
- To understand Common metrics of web as well as KPI related concepts.
- To learn about the various Web analytics versions.

UNIT I INTRODUCTION 9

Definition, Process, Key terms: Site references, Keywords and Key phrases; building block terms: Visit characterization terms, Content characterization terms, Conversion metrics; Categories: Offsite web, on site web; Web analytics platform, Web analytics evolution, Need for web analytics, Advantages, Limitations.

UNIT II DATA COLLECTION 9

Click stream Data: Web logs, Web Beacons, JavaScript tags, Packet Sniffing; Outcomes Data: E-commerce, Lead generation, Brand/Advocacy and Support; Research data: Mindset,

Organizational structure, Timing; Competitive Data: Panel-Based measurement, ISP-based measurement, Search Engine data.

UNIT III QUALITATIVE ANALYSIS 9

Heuristic evaluations: Conducting a heuristic evaluation, Benefits of heuristic evaluations; Site Visits: Conducting a site visit, Benefits of site visits; Surveys: Website surveys, Post-visit surveys, creating and running a survey, Benefits of surveys. Capturing data: Web logs or JavaScript's tags, Separate data serving and data capture, Type and size of data, Innovation, Integration, Selecting optimal web analytic tool, Understanding click stream data quality, Identifying unique page definition, Using cookies, Link coding issues.

UNIT IV WEB METRICS 9

Common metrics: Hits, Page views, Visits, Unique visitors, Unique page views, Bounce, Bounce rate, Page/visit, Average time on site, New visits; Optimization (e-commerce, non e-commerce sites): Improving bounce rates, Optimizing adwords campaigns; Real time report, Audience report, Traffic source report, Custom campaigns, Content report, Google analytics, Introduction to KPI, characteristics, Need for KPI, Perspective of KPI, Uses of KPI. Relevant Technologies: Internet & TCP/IP, Client / Server Computing, HTTP (Hypertext Transfer Protocol), Server Log Files & Cookies, Web Bugs.

UNIT V WEB ANALYTICS 2.0 9

Web analytics 1.0, Limitations of web analytics 1.0, Introduction to analytic 2.0, Competitive intelligence analysis : CI data sources, Toolbar data, Panel data ,ISP data, Search engine data, Hybrid data, Website traffic analysis: Comparing long term traffic trends, Analyzing competitive site overlap and opportunities. Google Analytics: Brief introduction and working, Adwords, Benchmarking, Categories of traffic: Organic traffic, Paid traffic; Google website optimizer, Implementation technology, Limitations, Performance concerns, Privacy issues**.**

TOTAL: 45 PERIODS

COURSE OUTCOMES

Upon completion of this course, the students should be able to:

CO1: Understand the Web analytics platform, and their evolution.
CO2: Use the various Data Streams Data.
CO3: Know how the survey of capturing of data will benefit.
CO4: Understand Common metrics of web as well as KPI related concepts.
CO5: Apply various Web analytics versions in existence.

REFERENCES:

1. Clifton B., Advanced Web Metrics with Google Analytics, Wiley Publishing, Inc.2nd ed, 2012.
2. Kaushik A., Web Analytics 2.0, The Art of Online Accountability and Science of Customer Centricity, Wiley Publishing, Inc. 1st ed, 2010.
3. Sterne J., Web Metrics: Proven methods for measuring web site success, John Wiley and Sons, 2002

MP4071 **HEALTHCARE ANALYTICS** **L T P C**
3 0 0 3

COURSE OBJECTIVES:

- To impart the fundamental concepts of Healthcare data analytics
- To give the knowledge about the Health care Data Sources.
- To familiarize Advanced Data Analytics for Healthcare
- To learn the Health IoT data analytics
- To implement the Applications and Practical Systems for Healthcare.

UNIT I INTRODUCTION 9

Introduction- Healthcare Data Sources and Basic Analytics - Healthcare Data Sources : Electronic Health Records: Components of HER- Coding system- Biomedical Image Analysis: Biomedical Imaging Modalities- Object Detection- Image Segmentation- Image Registration- Feature Extraction- Mining of Sensor Data in Healthcare: Mining Sensor Data in Medical Informatics: Scope and Challenges- Sensor Data Mining Applications

UNIT II HEALTHCARE DATA SOURCES 9

Biomedical Signal Analysis: Types of Biomedical Signals- ECG Signal Analysis- Denoising of Signals- Multivariate Biomedical Signal Analysis- Cross-Correlation Analysis- Methods to Study Connectivity- Genomic Data Analysis for Personalized Medicine: Genomic Data Generation- Methods and Standards for Genomic Data Analysis- Types of Computational Genomics Studies towards Personalized Medicine

UNIT III ADVANCED DATA ANALYTICS FOR HEALTHCARE 9

Basic Statistical Prediction Models- Alternative Clinical Prediction Models- Survival Models- Evaluation and Validation- Temporal Data Mining for Healthcare Data: Association Analysis- Temporal Pattern Mining- Sensor Data Analysis- Other Temporal Modeling Methods- Visual Analytics for Healthcare: Visual Analytics and Medical Data Visualization- Visual Analytics in Healthcare.

UNIT IV HEALTH IOT DATA ANALYTICS 9

Internet of things in the healthcare industry- IoT healthcare architecture- Characteristics of IoT health data- Health data analytics using Internet of things- Computational intelligence in Internet of things for future healthcare applications.

UNIT V APPLICATIONS AND PRACTICAL SYSTEMS FOR HEALTHCARE 9

Data Analytics for Pervasive Health: Supporting Infrastructure and Technology - Basic Analytic Techniques- Advanced Analytic Techniques- Applications - Fraud Detection in Healthcare- Data Analytics for Pharmaceutical Discoveries- Clinical Decision Support Systems.

COURSE OUTCOMES:

CO1: Describe the basics of healthcare data analytics.
CO2: Explain the Healthcare Data Sources.
CO3: Discuss the Advanced Data Analytics for Healthcare.
CO4: Express the Health IoT data analytics.
CO5: Apply the practical Systems for Healthcare.

TOTAL: 45 PERIODS

REFERENCES

1. Chandan K. Reddy , Charu C. Aggarwal, Healthcare Data Analytics 1st Edition, Kindle

Edition, CRC press, 2020.
2. Sanjay Kumar Singh Ravi Shankar Singh Anil Kumar Pandey Udmale S.S. Ankit Chaudhary , IoT-Based Data Analytics for the Healthcare Industry Techniques and Applications 1st Edition, Elsevier, Academic Press
3. Prasant Kumar Pattnaik, Suneeta Mohanty (Editor), Satarupa Mohanty (Editor) Format: Kindle Edition, Smart Healthcare Analytics in IoT Enabled Environment 1st edition Kindle Edition, Springer Nature Switzerland AG 2020
4. Nilanjan Dey, Amira Ashour, Simon James Fong , Chintan Bhatt , Healthcare Data Analytics and Management 1st Edition, Elsevier, Academic Press 2018.
5. Sanket Shah, Healthcare Analytics: A Comprehensive Guide, Kindly Edition, 2020

BD4008 STATISTICS FOR BUSINESS ANALYTICS L T P C
3 0 0 3

COURSE OBJECTIVES:

- To provide the required skill to apply the statistical tools in engineering Problems.
- To introduce the basic concepts of Time Series and Estimations.
- To acquaint the knowledge of Statistical Inference and Decision Theory.
- To provide the basic tools of Statistics for data analysis and Decision making on the sampling and inference.

UNIT I INTRODUCTION TO TIME SERIES 9

Time Series: Meaning And Need Of Time Series Analysis, Components Of Time Series, Additive And multiplicative Model, Utility Of Time Series. Methods Of Determining Trends. Components Of Timeseries, Smoothing Auto Correlation, Stationarity, Concepts Of AR, MA, ARMA & ARIMA Models with Illustrations.

UNIT II ESTIMATION 9

Methods of estimation: Random samples, sampling distributions of estimators, Methods of moments, Unbiasedness: Unbiased estimator, Illustration of unbiased estimator for the parameter and parametric function. Definitions of Consistency, Sufficient condition for consistency, concept of efficiency and sufficiency. Neyman- Factorization theorem (without proof), concept of likelihood function, Maximum Likelihood, Properties of MLE (without proof), Estimation of the parameters of normal distribution and other standard distributions by MLE.

UNIT III STATISTICAL INFERENCE AND DECISION THEORY 9

Statement and proof of Cramer Rao inequality. Definition of Minimum Variance Bound Unbiased Estimator (MVBUE) of $\varphi(\theta)$, (statement only). Rao-Blackwell theorem, Lehmann-Scheffe theorem. Definition of MVUE, Procedure to obtain MVUE (statement only), examples. Minimum Variance Unbiased Estimator (MVUE) and Uniformly Minimum Variance Unbiased Estimator(UMVUE). Basic elements of Statistical Decision Problem. Expected loss, decision rules(nonrandomized and randomized), decision principles (conditional Bayes, frequentist), inference as a decision problem, optimal decision rules. Bayes and minimax decision rule. Admissibility o fminimax rules and Bayes rules.

UNIT IV REGRESSION AND RELIABILITY 9

Multiple linear regression, forward, backward & stepwise regression, Logistic Regression. Reliability of system of independent components, association of random variables, bounds on system reliability, improved bounds on system reliability using modular decompositions. Replacement policy comparisons, preservation of life distribution classes under reliability operations. Reversed hazard rate, cumulative reversed hazard function, relation between hazard function and reversed hazard function. Reversed lack of memory property.

UNIT V STATISTICAL QUALITY CONTROL 9

Meaning and purpose of Statistical quality control, Concept of process control, product control, assignable causes, chance causes and rational subgroups. Control charts and their uses, Choice of subgroup sizes, Construction of control chart for (mean), R (range), s (standard deviation), c (no.of defectives), p (fraction defectives) with unequal subgroup size. Interpretation of non-random patterns of points. Modified control chart. CUSUM Chart. Consumer's risk, producer's risk, OCcurve, acceptance sampling plan by attributes and variables. Concept of Six Sigma.

TOTAL:45 PERIODS

COURSE OUTCOMES :

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

CO1: Perform time series analysis of data.
CO2: Apply the concept of Point estimation by Method of moments and Maximum likelihood estimation.
CO3: Evaluate the regression and reliability for the statistical sampling data.
CO4: Apply various estimators for the statistical concepts.
CO5: Apply various techniques in quality control and acceptance sampling.

REFERENCES:

1. Johnson, R.A., Miller, I and Freund J., "Miller and Freund‘s Probability and Statistics for Engineers", Pearson Education, Asia, 8th Edition, 2015.
2. Barlow, R. E. and Proschan F. (1996). Mathematical Theory of Reliability. John Wiley.
3. Statistical Inference: P. J. Bickel and K. A. Docksum, 2ndEdition, Prentice Hall
4. Duncan A.J. (1974): Quality Control and Industrial Statistics, IV Edition, Taraporewala and Sons.
5. M. Mitzenmacher and E. Upfal. Probability and Computing: Randomized Algorithms and Probabilistic Analysis. Cambridge University Press, 2005.
6. Devore. J.L., "Probability and Statistics for Engineering and the Sciences‖, Cengage Learning, New Delhi, 8th Edition, 2014.
7. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, “ Introduction to Linear Regression Analysis”,Wiley, 6th Edition, 2021.
8. Chris Chatfield “The Analysis of Time Series: An Introduction”, Chapman & Hall/CRC,Sixth Edition, 2003.
9. George Casella, Roger L. Berger, “Statistical Inference”, 2nd ed., Thomson Learning,2007.
10. Mukhopadhay, Parimal, ”Theory and Methods of Survey Sampling”, Prentice Hall,2008.
11. Tobias, P. A. and Trindane, D. C, “Applied Reliability”, Second edition, CRC Press,1995.
12. Rao, C.R.,” Linear Statistical Inference and its Applications”, Wiley Eastern,2009.

AUDIT COURSES

AX4091 ENGLISH FOR RESEARCH PAPER WRITING L T P C
2 0 0 0

COURSE OBJECTIVES:

- Teach how to improve writing skills and level of readability
- Tell about what to write in each section
- Summarize the skills needed when writing a Title
- Infer the skills needed when writing the Conclusion
- Ensure the quality of paper at very first-time submission

UNIT I INTRODUCTION TO RESEARCH PAPER WRITING 6

Planning and Preparation, Word Order, Breaking up long sentences, Structuring Paragraphs and Sentences, Being Concise and Removing Redundancy, Avoiding Ambiguity and Vagueness

UNIT II PRESENTATION SKILLS 6

Clarifying Who Did What, Highlighting Your Findings, Hedging and Criticizing, Paraphrasing and Plagiarism, Sections of a Paper, Abstracts, Introduction

UNIT III TITLE WRITING SKILLS 6

Key skills are needed when writing a Title, key skills are needed when writing an Abstract, key skills are needed when writing an Introduction, skills needed when writing a Review of the Literature, Methods, Results, Discussion, Conclusions, The Final Check

UNIT IV RESULT WRITING SKILLS 6

Skills are needed when writing the Methods, skills needed when writing the Results, skills are needed when writing the Discussion, skills are needed when writing the Conclusions

UNIT V VERIFICATION SKILLS 6

Useful phrases, checking Plagiarism, how to ensure paper is as good as it could possibly be the first- time submission

TOTAL: 30 PERIODS

COURSE OUTCOMES:

CO1 –Understand that how to improve your writing skills and level of readability
CO2 – Learn about what to write in each section
CO3 – Understand the skills needed when writing a Title
CO4 – Understand the skills needed when writing the Conclusion
CO5 – Ensure the good quality of paper at very first-time submission

REFERENCES:

1. Adrian Wallwork , English for Writing Research Papers, Springer New York Dordrecht Heidelberg London, 2011
2. Day R How to Write and Publish a Scientific Paper, Cambridge University Press 2006
3. Goldbort R Writing for Science, Yale University Press (available on Google Books) 2006
4. Highman N, Handbook of Writing for the Mathematical Sciences, SIAM. Highman's book 1998.

AX4092 **DISASTER MANAGEMENT** **L T P C**
2 0 0 0

COURSE OBJECTIVES:

- Summarize basics of disaster
- Explain a critical understanding of key concepts in disaster risk reduction and humanitarian response.
- Illustrate disaster risk reduction and humanitarian response policy and practice from multiple perspectives.
- Describe an understanding of standards of humanitarian response and practical relevance in specific types of disasters and conflict situations.
- Develop the strengths and weaknesses of disaster management approaches

UNIT I INTRODUCTION 6

Disaster: Definition, Factors and Significance; Difference between Hazard And Disaster; Natural and Manmade Disasters: Difference, Nature, Types and Magnitude.

UNIT II REPERCUSSIONS OF DISASTERS AND HAZARDS 6

Economic Damage, Loss of Human and Animal Life, Destruction Of Ecosystem. Natural Disasters: Earthquakes, Volcanisms, Cyclones, Tsunamis, Floods, Droughts And Famines, Landslides And Avalanches, Man-made disaster: Nuclear Reactor Meltdown, Industrial Accidents, Oil Slicks And Spills, Outbreaks Of Disease And Epidemics, War And Conflicts.

UNIT III DISASTER PRONE AREAS IN INDIA 6

Study of Seismic Zones; Areas Prone To Floods and Droughts, Landslides And Avalanches; Areas Prone To Cyclonic and Coastal Hazards with Special Reference To Tsunami; Post-Disaster Diseases and Epidemics

UNIT IV DISASTER PREPAREDNESS AND MANAGEMENT 6

Preparedness: Monitoring Of Phenomena Triggering a Disaster or Hazard; Evaluation of Risk: Application of Remote Sensing, Data from Meteorological And Other Agencies, Media Reports: Governmental and Community Preparedness.

UNIT V RISK ASSESSMENT 6

Disaster Risk: Concept and Elements, Disaster Risk Reduction, Global and National Disaster Risk Situation. Techniques of Risk Assessment, Global Co-Operation in Risk Assessment and Warning, People's Participation in Risk Assessment. Strategies for Survival

TOTAL : 30 PERIODS

COURSE OUTCOMES:

CO1: Ability to summarize basics of disaster

CO2: Ability to explain a critical understanding of key concepts in disaster risk reduction and humanitarian response.

CO3: Ability to illustrate disaster risk reduction and humanitarian response policy and practice from multiple perspectives.

CO4: Ability to describe an understanding of standards of humanitarian response and practical relevance in specific types of disasters and conflict situations.

CO5: Ability to develop the strengths and weaknesses of disaster management approaches

REFERENCES:

1. Goel S. L., Disaster Administration And Management Text And Case Studies",Deep & Deep Publication Pvt. Ltd., New Delhi,2009.
2. NishithaRai, Singh AK, "Disaster Management in India: Perspectives, issues and strategies "'NewRoyal book Company,2007.
3. Sahni, Pradeep Et.Al. ," Disaster Mitigation Experiences And Reflections", Prentice Hall OfIndia, New Delhi,2001.

AX4093 CONSTITUTION OF INDIA L T P C
2 0 0 0

COURSE OBJECTIVES:

Students will be able to:

- Understand the premises informing the twin themes of liberty and freedom from a civil rights perspective.
- To address the growth of Indian opinion regarding modern Indian intellectuals' constitutional
- Role and entitlement to civil and economic rights as well as the emergence of nationhood in the early years of Indian nationalism.
- To address the role of socialism in India after the commencement of the Bolshevik Revolution 1917 And its impact on the initial drafting of the Indian Constitution.

UNIT I HISTORY OF MAKING OF THE INDIAN CONSTITUTION

History, Drafting Committee, (Composition & Working)

UNIT II PHILOSOPHY OF THE INDIAN CONSTITUTION

Preamble, Salient Features

UNIT III CONTOURS OF CONSTITUTIONAL RIGHTS AND DUTIES

Fundamental Rights, Right to Equality, Right to Freedom, Right against Exploitation, Right to Freedom of Religion, Cultural and Educational Rights, Right to Constitutional Remedies, Directive Principles of State Policy, Fundamental Duties.

UNIT IV ORGANS OF GOVERNANCE

Parliament, Composition, Qualifications and Disqualifications, Powers and Functions, Executive, President, Governor, Council of Ministers, Judiciary, Appointment and Transfer of Judges, Qualifications, Powers and Functions.

UNIT V LOCAL ADMINISTRATION

District's Administration head: Role and Importance, Municipalities: Introduction, Mayor and role of Elected Representative, CEO, Municipal Corporation. Pachayati raj: Introduction, PRI: Zila Panchayat. Elected officials and their roles, CEO Zila Pachayat: Position and role. Block level: Organizational Hierarchy(Different departments), Village level:Role of Elected and Appointed officials, Importance of grass root democracy.

UNIT VI ELECTION COMMISSION

Election Commission: Role and Functioning. Chief Election Commissioner and Election Commissioners - Institute and Bodies for the welfare of SC/ST/OBC and women.

TOTAL: 30 PERIODS

COURSE OUTCOMES:

Students will be able to:

1. Discuss the growth of the demand for civil rights in India for the bulk of Indians before the arrival of Gandhi in Indian politics.
2. Discuss the intellectual origins of the framework of argument that informed the conceptualization of social reforms leading to revolution in India.
3. Discuss the circumstances surrounding the foundation of the Congress Socialist Party[CSP] under the leadership of Jawaharlal Nehru and the eventual failure of the proposal of direct elections through adult suffrage in the Indian Constitution.
4. Discuss the passage of the Hindu Code Bill of 1956.

SUGGESTED READING

1. The Constitution of India,1950(Bare Act),Government Publication.
2. Dr.S.N.Busi, Dr.B. R.Ambedkar framing of Indian Constitution,1st Edition, 2015.
3. M.P. Jain, Indian Constitution Law, 7th Edn., LexisNexis,2014.
4. D.D. Basu, Introduction to the Constitution of India, LexisNexis, 2015.

AX494 **நற்றமிழ் இலக்கியம்** **L T P C**

2 0 0 0

UNIT I **சங்க இலக்கியம்** **6**

1.தமிழின் துவக்க நூல் தொல்காப்பியம்
– எழுத்து, சொல், பொருள்

2.அகநானூறு (82)
- இயற்கை இன்னிசை அரங்கம்

3.குறிஞ்சிப் பாட்டின் மலர்க்காட்சி

4.புறநானூறு (95,195)
- போரை நிறுத்திய ஔவையார்

UNIT II **அறநெறித் தமிழ்** **6**

1. அறநெறி வகுத்த திருவள்ளுவர்
- அறம் வலியுறுத்தல், அன்புடைமை, ஒப்புறவு அறிதல், ஈகை, புகழ்
2. பிற அறநூல்கள் - இலக்கிய மருந்து
– ஏலாதி, சிறுபஞ்சமூலம், திரிகடுகம், ஆசாரக்கோவை (தூய்மையை வலியுறுத்தும் நூல்)

UNIT III **இரட்டைக் காப்பியங்கள்** **6**

1. கண்ணகியின் புரட்சி
- சிலப்பதிகார வழக்குரை காதை
2. சமூகசேவை இலக்கியம் மணிமேகலை
- சிறைக்கோட்டம் அறக்கோட்டமாகிய காதை

UNIT IV **அருள்நெறித் தமிழ்** **6**

1. சிறுபாணாற்றுப்படை
- பாரி முல்லைக்குத் தேர் கொடுத்தது, பேகன் மயிலுக்குப்

போர்வை கொடுத்தது, அதியமான் ஔவைக்கு நெல்லிக்கனி கொடுத்தது, அரசர் பண்புகள்

2. நற்றிணை
   - அன்னைக்குரிய புன்னை சிறப்பு
3. திருமந்திரம் (617, 618)
   - இயமம் நியமம் விதிகள்
4. தர்மச்சாலையை நிறுவிய வள்ளலார்
5. புறநானூறு
   - சிறுவனே வள்ளலானான்
6. அகநானூறு (4) - வண்டு
   நற்றிணை (11) - நண்டு
   கலித்தொகை (11) - யானை, புறா
   ஐந்திணை 50 (27) - மான்

ஆகியவை பற்றிய செய்திகள்

UNIT V **நவீன தமிழ் இலக்கியம்** **6**

1. உரைநடைத் தமிழ்,
   - தமிழின் முதல் புதினம்,
   - தமிழின் முதல் சிறுகதை,
   - கட்டுரை இலக்கியம்,
   - பயண இலக்கியம்,
   - நாடகம்,
2. நாட்டு விடுதலை போராட்டமும் தமிழ் இலக்கியமும்,
3. சமுதாய விடுதலையும் தமிழ் இலக்கியமும்,
4. பெண் விடுதலையும் விளிம்பு நிலையினரின் மேம்பாட்டில் தமிழ் இலக்கியமும்,
5. அறிவியல் தமிழ்,
6. இணையத்தில் தமிழ்,
7. சுற்றுச்சூழல் மேம்பாட்டில் தமிழ் இலக்கியம்.

TOTAL: 30 PERIODS

<u>தமிழ் இலக்கிய வெளியீடுகள் / புத்தகங்கள்</u>

1. தமிழ் இணைய கல்விக்கழகம் (Tamil Virtual University)
- www.tamilvu.org
2. தமிழ் விக்கிப்பீடியா (Tamil Wikipedia)
-https://ta.wikipedia.org
3. தர்மபுர ஆதீன வெளியீடு
4. வாழ்வியல் களஞ்சியம்
- தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்
5. தமிழ்கலைக் களஞ்சியம்
- தமிழ் வளர்ச்சித் துறை (thamilvalarchithurai.com)
6. அறிவியல் களஞ்சியம்
- தமிழ்ப் பல்கலைக்கழகம், தஞ்சாவூர்