GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT

(GITAM)

(Deemed to be University, Estd. u/s 3 of UGC Act 1956)

⁎VISAKHAPATNAM ⁎HYDERABAD ⁎BENGALURU⁎

Accredited by NAAC with 'A' Grade

REGULATIONS & SYLLABUS

Master of Technology

in

Data Science

(W.e.f 2017-18 admitted batch)

Website: www.gitam.edu

Semester –I

S.No	Course Code	Course Title	Category	L	T	P	C
1	**EIT705**	Linear algebra & Probability Theory	**CE**	**4**	**0**	**0**	**4**
2	**EIT707**	Machine Learning	**CE**	**4**	**0**	**0**	**4**
3	**ECS701**	Advanced Data Structures and Algorithms	**CE**	**4**	**0**	**0**	**4**
4	**EITXXX**	Program Elective I	**PE**	**3**	**0**	**0**	**3**
5	**EITXXX**	Program Elective II	**PE**	**3**	**0**	**0**	**3**
6	**EYYXXX**	Interdisciplinary Elective - I	**IDE**	**3**	**0**	**0**	**3**
7	**EIT721**	Advanced data structures Lab	**CE**	**0**	**0**	**3**	**2**
8	**ECS723**	Data Analytics Lab	**CE**	**0**	**0**	**3**	**2**
							25

Semester –II

S.No	Course Code	Course Title	Category	L	T	P	C
1	**EIT706**	NoSQL Databases	**CE**	**4**	**0**	**0**	**4**
2	**EIT708**	Big Data and Analytics	**CE**	**4**	**0**	**0**	**4**
3	**EIT710**	Social Network analysis	**CE**	**4**	**0**	**0**	**4**
4	**EITXXX**	Program Elective-III	**PE**	**3**	**0**	**0**	**3**
5	**EITXXX**	Program Elective-IV	**PE**	**3**	**0**	**0**	**3**
6	**EYY XXX**	Interdisciplinary Elective - II	**IDE**	**3**	**0**	**0**	**3**
7	**EIT726**	Hadoop Eco System Lab	**CE**	**0**	**0**	**3**	**2**
8	**EIT728**	Big Data Analytics Lab	**CE**	**0**	**0**	**3**	**2**
9	**EIT792**	Seminar	**CE**			**3**	**2**
							27

Semester –III

S.No.	Course Code	Course Title	Category	L	T	P	C
1	**EIT891**	Project Work-I	**PW**				**8**
2	**EIT893**	Comprehensive Viva-voce	**CE**				**2**
							10

Semester- IV

S.No.	Course Code	Course Title	Category	L	T	P	C
1	**EIT892**	Project work-II	**PW**				**14**
							14

Interdisciplinary Elective-I

S.No.	Course Code	Course Title	Category	L	T	P	C
1	EID774	Fraud analytics	IDE	3	0	0	3
2	EID772	E-commerce	IDE	3	0	0	3
3	EIT741	Information Storage management	IDE	3	0	0	3

Interdisciplinary Elective-II

S.No.	Course Code	Course Title	Category	L	T	P	C
1	EIT744	Digital Image Processing	IDE	3	0	0	3
2	ECS749	Internet of Things	IDE	3	0	0	3
3	EID777	Graph theory and Algorithms	IDE	3	0	0	3

Program Elective-I

S.No	Course Code	Course Title	Category	L	T	P	C
1	EIT749	Optimization Techniques	PE	3	0	0	3
2	EIT751	Exploratory Data analysis	PE	3	0	0	3
3	EIT753	Game Theory	PE	3	0	0	3

Program Elective-II

S.No	Course Code	Course Title	Category	L	T	P	C
1	ECS746	Natural Language Processing	PE	3	0	0	3
2	EIT755	Web Semantics	PE	3	0	0	3
3	ECS748	Cloud computing	PE	3	0	0	3

Program Elective-III

S.No.	Course Code	Course Title	Category	L	T	P	C
1	EIT752	Stream Mining	PE	3	0	0	3
2	EIT742	Bioinformatics	PE	3	0	0	3
3	EIT754	Mining Massive Data Sets	PE	3	0	0	3

Program Elective-IV

S.No.	Course Code	Course Title	Category	L	T	P	C
1	EIT756	Information Retrieval	PE	3	0	0	3
2	EIT758	Deep Learning	PE	3	0	0	3
3	EIT760	Multimedia Data analytics	PE	3	0	0	3

Number of Credits

Semester	I	II	III	IV	Total
Credits	25	27	10	14	76

EIT705 LINEAR ALGEBRA & PROBABILITY THEORY

L T P C

4 0 0 4

Module I **10 hours**

Solving Linear Equations : Introduction to Vectors , Vectors and Linear Equations, The Idea of Elimination, Elimination Using Matrices, Rules for Matrix Operations, Inverse Matrices , Elimination = Factorization: $A = LU$, Transposes and Permutations. Vector Spaces and Subspaces : Spaces of Vectors, The Nullspace of A: Solving $Ax = 0$ and $Rx = 0$, The Complete Solution to $Ax = b$, Dimensions of the Four Subspaces

Module II **8 hours**

Orthogonality : Orthogonality of the Four Subspaces, Projections, Least Squares Approximations, Orthonormal Bases and Gram-Schmidt. **Determinants** : The Properties of Determinants, Permutations and Cofactors, Cramer's Rule, Inverses and Volumes.

Module III **8 hours**

Eigen values and Eigen vectors : Introduction to Eigenvalues, Diagonalizing a Matrix, Systems of Differential Equations, Symmetric Matrices, Positive Definite Matrices. Singular Value Decomposition (SVD) ,Applications.

Module IV **10 hours**

Probability distributions:- Introduction to probability and random variables - Binomial distribution, Poisson distribution, Geometric distribution, Normal distribution, Log-Normal distribution, Gamma distribution, Beta distribution & Weibull distribution - Random samples and sampling distributions of mean and variance.

Module V **8 hours**

Parameter Estimation: Point estimation - Properties of estimators, The method of maximum likelihood, The method of moments, Confidence interval estimation of mean, and variance.

Statistical hypothesis tests: Operations characteristic curve, Tests of hypothesis on the mean of a Normal Distribution, Tests of hypothesis on the means of two Normal distributions, The paired t-test,Tests of hypothesis on one variance, Tests of hypothesis for the equality of two variances, The testing for goodness of fit.

Text Book:

1. Gilbert Strang, Introduction to Linear Algebra, 5th Edn. Wellesley-Cambridge Press, 2016.
2. Miller & Freund's Probability and Statistics for Engineers, Prentice Hall, 2011.

EIT707 MACHINE LEARNING

L T P C

4 0 0 4

Module I **8 hours**

Introduction, Maximum likelihood estimation, linear regression, least squares, geometric view, ridge regression, probabilistic views of linear regression

Module II **8 hours**

Model Assessment and Selection: bias-variance, Bayes rule, maximum a posteriori, The Gaussian Distribution

Module III **10 hours**

Bayesian linear regression, sparsity, subset selection for linear regression, Subset Selection, Shrinkage Methods, Methods Using Derived Input Directions, A Comparison of the Selection and Shrinkage Methods, Lasso and Related Path Algorithms

Module IV **9 hours**

Logistic and probit classifiers, kernel methods, Gaussian processes, maximum margin, support vector machines, trees, random forests, boosting

Module V **10 hours**

Clustering, k-means, EM algorithm, missing data, mixtures of Gaussians, matrix factorization, PCA and variations, Markov models, hidden Markov models, model selection

Text Book(s)

1. T.Hastie,R.Tibshirani and J.Friedman, The Elements of Statistical Learning, 2/e, Springer.

2. C. Bishop, Pattern Recognition and Machine Learning, Springer.

References

1. Tom M. Mitchell, Machine Learning, McGraw Hill, 2013.

2. M Narasimha Murty, Introduction to Pattern Recognition and Machine Learning, World Scientific Publishing Company, 2015

ECS701 ADVANCED DATA STRUCTURES & ALGORITHMS

L T P C

4 0 0 4

Module I 12hrs

Introduction to Data Structures and Algorithms, Performance Analysis: Time Complexity, Space Complexity, Amortized Complexity, Asymptotic Notations, Randomized Algorithms, Linked List, Stacks, Queues, Sparse Matrices. Algebraic Problems: General Method, Evaluation and Interpolation.

Module II 10hrs

Introduction to Graphs, Graph Traversal. Introduction to Trees and Tree Traversals, Binary Search Trees, AVL Trees, B-Trees, Priority Queues.

Module III 10hrs

Divide and Conquer: General Method, Selection Problem, Strassen's Matrix Multiplication, and Convex Hull Problem. The Greedy Method: General Method, Knapsack, Job Sequencing with Dead Lines, Minimum Cost Spanning Trees using Kruskal's Algorithm, using UNION and FIND, Dijkstra's algorithm for single source shortest path.

Module IV 10hrs

Dynamic Programming: General Method, Matrix Chain Multiplication, Longest Common Subsequence, Reliability Design, Traveling Sales Person Problem. Back Tracking: General Method, 8 Queens Problem, Hamiltonian Cycle, Graph Coloring Problem.

Module V 10hrs

Branch-and-Bound: General Method, FIFO Branch and Bound, LIFO Branch and Bound, LC Branch and Bound, Traveling Sales Person Problem. P-Class Problem, NP-Class Problems, NP-Complete Problems, NP-Hard problems.

Text Book(s)

1. Ellis Horowitz, Sartaz Sahni, Sanguthevar Rajasekharan , Fundamentals of Computer Algorithms , II Edition, University Press.

2. Sartaj Sahni, Data Structures, Algorithms and Applications in C++, 2/e, Universities Press.

3. Varsha H Patil, Data Structures using C++, Oxford Higher Education.

References

1. Thomas H. Cormen, et al., Introduction to Algorithms,3/e, MIT Press.

2. Mark Allen Weiss, Data Structures and Algorithms.

3. Adam Drozdek, Data Structures and Algorithms in C++, 3/e, Cengage Learning.

4. Michel T. Goddrich,Roberto Tamassia, Algorithm Design John Weily and Sons

5. Web Resources: http://www.personal.kent.edu/~rmuhamma/Algorithms/ algorithm.html

EIT749 OPTIMIZATION TECHNIQUES

(Program Elective I)

L	T	P	C
3	0	0	3

Module I **8 hours**

Introduction to Optimization: Introduction, engineering applications of optimization, statement of an optimization problem-design vector, design constraints, constraint surface, objective function, classification of optimization problems, optimization techniques.

Module II **8 hours**

Classical optimization techniques: Introduction, single variable optimization algorithms, Optimality Criteria, Bracketing Methods, Region Elimination Methods, Point Elimination Methods, Gradient Elimination Methods.

Module III **8 hours**

Multi variable optimization algorithms: Optimality Criteria, Unidirectional Search, Direct Search Methods, Gradient based Methods. Constrained Optimization algorithms: Kuhn – Tucker Conditions, Lagrangian Duality Theory, Transformation Methods, Sensitivity Analysis, Direct Search and Linear Search Techniques, Quadratic Programming, Gradient Projection method.

Module IV **8 hours**

Evolutionary Algorithms: Introduction to genetic algorithms(GA), Binary GAs, Real Parameter GAs, Evolution Strategies, Evolutionary Programming, Genetic Programming, Multi-Model Function Optimization.

Module V **10 hours**

Multi –Objective Evolutionary Algorithms: Non Elitist - Vector Evaluated GA, Weight Based GA, Multi-objective GA, Non Dominated Sorting GA, Niched Pareto GA, Predator- Prey Evolution Strategy. Elitist- Rudolph's Elitist Algorithm, Elitist Non dominated Sorting GA, Distance based Pareto GAs, Constrained Algorithms – Penalty approach, Jimenes-Verdagay-Gomes –Skarmeta's method, Constrained Tournament method, Ray-Tai-Seow's method.

Text Book(s)

1. Singiresu S. Rao, Engineering Optimization, theory and practice, New Age International Pvt. Ltd., 3rd Edition, 2013.

2. Kalyan Moy Deb., Optimization for Engineering Design, 2/e, Prentice Hall of India, 2012.
3. Kalyan Moy Deb., Multi-objective optimization using evolutionary algorithms, John Wiley & Sons,1st Edition, 2001

References

1. David Edward Goldberg, Genetic algorithms in search, optimization, and machine learning, Addison-Wesley Publishing ,1989
2. Gupta P K., & Hira D.S., Operation Research, 6/e, S Chand Publishers, 2006.

EIT751 EXPLORATORY DATA ANALYSIS

(Program Elective I)

L	T	P	C
3	0	0	3

Module I **9 hours**

Managing Data Frames with the dplyr package: Data Frames, the dplyr Package, dplyr Grammar, Installing the dplyr package, select(), ilter(), arrange(), rename(), mutate(), group_by(), Exploratory Data Analysis : Formulate your question, Read in your data, Check the packaging, Run str().

Module II **10 hours**

Principles of Analytic Graphics: Show comparisons, Show causality, mechanism, explanation, systematic structure, Show multivariate data, Integrate evidence, Describe and document the evidence, Exploratory Graphs: Characteristics of exploratory graphs, Air Pollution in the United States, Getting the Data, Simple Summaries: One Dimension, Five Number Summary, Boxplot, Histogram, Overlaying Features, Barplot, Simple Summaries: Two Dimensions and Beyond, Multiple Boxplots, Multiple Histograms, Scatterplots, Scatterplot - Using Color, Multiple Scatterplots

Module III **9 hours** Plotting Systems: The Base Plotting System, The Lattice System, The ggplot2 System

Graphics Devices: The Process of Making a Plot, How Does a Plot Get Created?, Graphics File Devices, Multiple Open Graphics Devices, Copying Plots

The Base Plotting System: Base Graphics, Simple Base Graphics, Some Important Base Graphics Parameters, Base Plotting Functions, Base Plot with Regression Line, Multiple Base Plots

Plotting and Color in R: Colors 1, 2, and 3, Connecting colors with data, Color Utilities in R, colorRamp(), colorRampPalette(), RColorBrewer Package, Using the RColorBrewer palettes, The smoothScatter() function, Adding transparency

Module IV **10 hours**

Hierarchical Clustering: Hierarchical clustering, How do we define close?, Example: Euclidean

distance, Example: Manhattan distance, Example: Hierarchical clustering, Prettier dendrograms, Merging points: Complete, Merging points: Average, Using the heatmap() function,

K-Means Clustering: Illustrating the K-means algorithm, Stopping the algorithm, Using the kmeans() function, Building heatmaps from K-means solutions, Notes and further resources

Dimension Reduction: Matrix data, Patterns in rows and columns, Related problem, SVD and PCA, Unpacking the SVD: u and v, SVD for data compression, Components of the SVD - Variance explained, Relationship to principal components, What if we add a second pattern?, Dealing with missing values.

Module V **9 hours**

The ggplot2 Plotting System: Part 1

The Basics: qplot(), Before You Start: Label Your Data, ggplot2 “Hello, world!”, Modifying aesthetics, Adding a geom, Histograms, Facets, Case Study: MAACS Cohort, Summary of qplot().

Text Book(s)

1.Roger D.Peng, Exploratory Data Analysis with R.

References

1.John W. Tukey, Exploratory Data Analysis, Addison-Wesley

EIT753 GAME THEORY

(Program Elective I)

L	T	P	C
3	0	0	3

Module I **10 hours**

Introduction of Game Theory Games and Solutions, Game Theory and the Theory of Competitive Equilibrium, Rational Behavior, the Steady State and Deductive Interpretations, Bounded Rationality, Terminology and Notation. Strategic Games : Nash Equilibrium, Strategic Games, Nash Equilibrium, examples, Existence of a Nash Equilibrium, Strictly Competitive Games, Bayesian Games: Strategic Games with Imperfect Information, Mixed, Correlated, and Evolutionary Equilibrium, Mixed Strategy Nash Equilibrium, Interpretations of Mixed Strategy Nash Equilibrium, Correlated Equilibrium, Evolutionary Equilibrium.

Module II **10 hours**

Extensive Games with Perfect Information, Extensive Games with Perfect Information, Sub-game Perfect Equilibrium, Two Extensions of the Definition of a Game, The Interpretation of a Strategy, Two Notable Finite Horizon Games, Iterated Elimination of Weakly Dominated Strategies.

Module III **10 hours**

Introduction of Extensive Games with Imperfect Information, Principles for the Equivalence of Extensive Games, Framing effects and the Equivalence of Extensive Games, Mixed and Behavioral Strategies Nash Equilibrium.

Module IV **10 hours**

Strategies and Beliefs, Sequential Equilibrium, Games with Observable Actions: Perfect Bayesian, Equilibrium, Refinements of Sequential Equilibrium, Trembling Hand Perfect Equilibrium.

Module V **10 hours**

The Core: Coalitional Games with Transferable Payo, Nonemptiness of the Core, Markets with Transferable Payo, Coalitional Games without Transferable Payo, Exchange Economies.

Text book:

1. An Introduction to Game Theory by Martin J.Osborne, MIT Press

ECS746 NATURAL LANGUAGE PROCESSING

(Program Elective II)

L T P C

3 0 0 3

Module I **12 hours**

Introduction and Overview What is Natural Language Processing, hands-on demonstrations. Ambiguity and uncertainty in language. The Turing test. **Regular Expressions** Chomsky hierarchy, regular languages, and their limitations. Finite-state automata. Practical regular expressions for finding and counting language phenomena. A little morphology. Exploring a large corpus with regex tools.

Programming in Python An introduction to programming in Python. Variables, numbers, strings, arrays, dictionaries, conditionals, iteration. The NLTK (Natural Language Toolkit) **String Edit Distance and Alignment** Key algorithmic tool: dynamic programming, a simple example, use in optimal alignment of sequences. String edit operations, edit distance, and examples of use in spelling correction, and machine translation.

Module II **10 hours**

Context Free Grammars Constituency, CFG definition, use and limitations. Chomsky Normal Form. Top-down parsing, bottom-up parsing, and the problems with each. The desirability of combining evidence from both directions **Non-probabilistic Parsing** Efficient CFG parsing with CYK, another dynamic programming algorithms. Early parser. Designing a little grammar, and parsing with it on some test data. **Probability** Introduction to probability theory Joint and conditional probability, marginals, independence, Bayes rule, combining evidence. Examples of applications in natural language. **Information Theory** The "Shannon game"--motivated by language! Entropy, crossentropy, information gain. Its application to some language phenomena.

Module III **12 hours**

Language modeling and Naive Bayes Probabilistic language modeling and its applications. Markov models. N-grams. Estimating the probability of a word, and smoothing. Generative models of language. Part of Speech Tagging and Hidden Markov Models, Viterbi Algorithm for Finding Most Likely HMM Path Dynamic programming with Hidden Markov Models, and its use for part-of-speech tagging, Chinese word segmentation, prosody, information extraction, etc.

Module IV **10 hours**

Probabilistic Context Free Grammars Weighted context free grammars. Weighted CYK. Pruning and beam search. **Parsing with PCFGs** A tree bank and what it takes to create one. The probabilistic version of CYK. Also: How do humans parse? Experiments with eye-tracking. Modern parsers. **Maximum Entropy Classifiers** The maximum entropy principle and its relation to maximum likelihood. Maximum entropy classifiers and their application to document classification, sentence segmentation, and other language tasks

Module V **10 hours**

Maximum Entropy Markov Models & Conditional Random Fields Part-of-speech tagging, noun-phrase segmentation and information extraction models that combine maximum entropy and finite-state machines. State-of-the-art models for NLP. **Lexical Semantics** Mathematics of Multinomial and Dirichlet distributions, Dirichlet as a smoothing for multinomial's.**Information Extraction** & Reference Resolution- Various methods, including HMMs. Models of anaphora resolution. Machine learning methods for co reference.

Text Books:

1. "Speech and Language Processing": Jurafsky and Martin, Prentice Hall
2. "Statistical Natural Language Processing"- Manning and Schutze, MIT Press
3. "Natural Language Understanding". James Allen. The Benajmins/Cummings Publishing Company

References:

1. Cover, T. M. and J. A. Thomas: Elements of Information Theory. Wiley.
2. Charniak, E.: Statistical Language Learning. The MIT Press.
3. Jelinek, F.: Statistical Methods for Speech Recognition. The MIT Press.
4. Lutz and Ascher - "Learning Python", O'Reilly

EIT755 WEB SEMENTICS

(Program Elective II)

L T P C

3 0 0 3

Module I **10 hours**

Web Intelligence: Thinking and intelligent web applications, the information age ,the world wide web, limitations of today's web, the next generation web, machine intelligence, artificial intelligence, ontology, inference engines, software agents, berners-lee www, semantic road map, logic on the semantic web.

Module II **12 hours**

Knowledge Representation for the Semantic Web: Ontologies and their role in the semantic web, ontologies languages for the semantic web – Resource Description Framework(RDF) / RDF schema, Ontology Web Language(OWL),UML,XML/XML schema.

Module III **8 hours**

Ontology Engineering: Ontology engineering, constructing ontology, ontology development tools, ontology methods, ontology sharing and merging, ontology libraries and ontology mapping, logic, rule and inference engines.

Module IV **10 hours**

Semantic Web Applications, Services and Technology: Semantic web applications and services, semantic search, e-learning, semantic bioinformatics, knowledge base, XML based web services, creating an OWL-S ontology for web services, semantic search technology, web search agents and semantic methods.

Module V **12 hours**

Social Network Analysis and semantic web: What is social networks analysis, development of the social networks analysis, electronic sources for network analysis, electronic discussion networks, blogs and online communities, web based networks, building semantic web applications with social network features.

Text Book(s)

1. Berners Lee, Godel and Turing, Thinking on the Web, Wiley Inderscience, 2008.

2. Peter Mika, Social Networks and the Semantic Web, Springer, 2007.

References

1. Davies, R.Studer, P.Warren, Semantic Web Technologies, Trends and Research in Ontology Based Systems, John Wiley & Sons, 2007.

2. Liyang Lu Chapman and Hall, Semantic Web and Semantic Web Services, CRC Press, 2007.

3. Heiner Stuckenschmidt, Frank Van Harmelen, Information Sharing on the semantic Web, Springer, 2004.

4. T.Segaran, C.Evans, J.Taylor , Programming the Semantic Web, O'Reilly,2009.

ECS748 CLOUD COMPUTING

(Program Elective II)

L T P C

3 0 0 3

Module I **8 hours**

Understanding Cloud Computing: Cloud origins and influences, basic concepts and terminology, goals and benefits, risks and challenges. **Fundamental Concepts and Models:** Roles and boundaries, cloud characteristics, cloud delivery models, cloud deployment models.

Module II **8 hours**

Cloud Enabling Technology: Data center technology, virtualization technology, web technology, multitenant technology, service technology.

Module III **9 hours**

Cloud Infrastructure Mechanisms: Logical network perimeter, virtual server, cloud storage device, cloud usage monitor, resource replication.

Module IV **10 hours**

Fundamental Cloud Architectures: Workload distribution architecture, resource pooling architecture, dynamic scalability architecture, elastic resource capacity architecture, service load balancing architecture, cloud bursting architecture, elastic disk provisioning architecture, redundant storage architecture.

Module V **9 hours**

Cloud Delivery Model Considerations: The cloud provider perspective- Building IaaS environments, equipping PaaS environments, optimizing SaaS environments, the cloud consumer perspective, working with IaaS environments, working with PaaS environments, working with SaaS services.

Text Book(s)

1.Thomas Erl, Ricardo Puttini, Zaigham Mahmood, Cloud Computing: Concepts, Technology & Architecture, PHI Publications, 2013.

References

1. John W. Rittinghouse, James F.Ransome, Cloud Computing: Implementation, Management and Security, CRC Press, 2012.
2. Anthony T.Velte, Toby J Velte Robert Elsenpeter, Cloud Computing –A Practical Approach, TMH Publications, 2010.
3. Michael Miller, Cloud Computing: Web Based Applications That Change the Way You Work and Collaborate Online, Que Publishing, 2008.

EID774 FRAUD ANALYTICS

(Interdisciplinary Elective I)

L	T	P	C
3	0	0	3

Module 1 **10 Hours**

Fraud: Detection, Prevention, and Analytics Fraud Detection and Prevention, Big Data for Fraud Detection, Data-Driven Fraud Detection, Fraud-Detection Techniques, Fraud Cycle, The Fraud Analytics Process Model, Scientific Perspective on Fraud

Module II **10 Hours**

Data Collection, Sampling, and Preprocessing Types of Data Sources, Merging Data Sources, Sampling, Types of Data Elements, Visual Data Exploration and Exploratory Statistical Analysis, Benford's Law, Descriptive Statistics, Missing Values, Outlier Detection and Treatment, Red Flags, Standardizing Data, Categorization, Weights of Evidence Coding, Variable Selection, Principal Components Analysis, RIDITs, PRIDIT Analysis, Segmentation

Module III **10 Hours**

Descriptive Analytics for Fraud Detection Graphical Outlier Detection Procedures, Statistical Outlier Detection Procedures, Break-Point Analysis, Peer-Group Analysis, Association Rule Analysis, Clustering, Introduction to Distance Metrics, Hierarchical Clustering, Example of Hierarchical Clustering Procedures, k-Means Clustering, Self-Organizing Maps, Clustering with Constraints, Evaluating and Interpreting Clustering Solutions, One-Class SVMs

Module IV **10 Hours**

Predictive Analytics for Fraud Detection Target Definition, Linear Regression, Logistic Regression, Basic Concepts, Logistic Regression Properties, Building a Logistic Regression Scorecard, Variable Selection for Linear and Logistic Regression, Decision Trees, Basic Concepts, Splitting Decision, Stopping Decision, Decision Tree Properties, Regression Trees, Using Decision Trees in Fraud Analytics,

Module V **10 Hours**

Social Network Analysis for Fraud Detection Homophily, Impact of the Neighborhood: Metrics, Neighborhood Metrics, Centrality Metrics, Collective Inference Algorithms, Featurization: Summary Overview, Community Mining: Finding Groups of Fraudsters, Extending the Graph: Toward a Bipartite Representation, Multipartite Graphs

Text Book:

1. Bart Baesens, Véronique Van Vlasselaer, Wouter Verbeke - FRAUD ANALYTICS Using Descriptive, Predictive, and Social Network Techniques(A Guide to Data Science for Fraud Detection), Wiley - August 2015

EID 772 E-COMMERCE

(Interdisciplinary Elective I)

L T P C

3 0 0 3

Module I **10 Hours**

Electronic commerce, Frame work, anatomy of e-commerce applications, e- commerce consumer applications, e-commerce organization applications, consumer oriented electronic commerce , mercantile process models.

Module II **10 Hours**

Electronic payment systems, digital token based smart cards, credit cards, risks in electronic payment systems, inter organizational commerce, EDI, EDI implementation, value added networks.

Module III **10 Hours**

Intra organizational commerce, work flow, automation customization and internal commerce, supply chain management, corporate digital library, document library, digital document types, corporate data warehouses, advertising and marketing, information based marketing, advertising on internet, online marketing process, market research.

Module IV **10 Hours**

Consumer search and resource discovery , information search and retrieval, commerce catalogues, information filtering.

Module V **10 Hours**

Multimedia, key multimedia concepts, digital video and electronic commerce, desktop video processing, desktop video conferencing.

Text Book(s)

1. Kalakata, Whinston, Frontiers of electronic commerce , Pearson Education,1996.

References

1. Hendry Chan, Raymond Lee, Tharam Dillon, Ellizabeth Chang, E-Commerce Fundamentals and Applications, Wiley, 2001.

2. Efrain Turbon, Jae Lee, David King, H. Michael Chang,E-Commerce,3/e, Pearson

Education, 2011.

3. Kenneth, Taudon, Carol Guyerico Traver, E-Commerce – Business, Technology, Society, 11/e, Prentice Hall of India, 2014.

EIT741 INFORMATION STORAGE MANAGEMENT

(Interdisciplinary Elective I)

L T P C

3 0 0 3

MODULE I **9 hours**

INTRODUCTION TO STORAGE AND MANAGEMENT

Introduction to Information Storage Management - Data Center Environment– Database Management System (DBMS) - Host - Connectivity –Storage-Disk Drive Components- Intelligent Storage System -Components of an Intelligent Storage System- Storage Provisioning- Types of Intelligent Storage Systems.

MODULE II **9 hours**

STORAGE NETWORKING

Fibre Channel: Overview - SAN and Its Evolution -Components of FC SAN -FC Connectivity-FC Architecture- IPSAN-FCOE-FCIP-Network-Attached Storage- General-Purpose Servers versus NAS Devices - Benefits of NAS- File Systems and Network File Sharing-Components of NAS - NAS I/O Operation -NAS Implementations -NAS File-Sharing Protocols-Object-Based Storage Devices-Content-Addressed Storage -CAS Use Cases.

MODULE III **9 hours**

BACKUP AND RECOVERY

Business Continuity -Information Availability -BC Terminology-BC Planning Life Cycle – Failure Analysis-Business Impact Analysis-Backup and Archive-Backup Purpose-Backup Considerations-Backup Granularity-Recovery Considerations-Backup Methods-Backup Architecture - Backup and Restore Operations.

MODULE IV **9 hours**

CLOUD COMPUTING

Cloud Enabling Technologies -Characteristics of Cloud Computing -Benefits of Cloud Computing - Cloud Service Models-Cloud Deployment models-Cloud computing Infrastructure-Cloud Challenges.

MODULE V **9 hours**

SECURING AND MANAGING STORAGE INFRASTRUCTURE

Information Security Framework -Storage Security Domains-Security Implementations in Storage Networking - Monitoring the Storage Infrastructure -Storage Infrastructure Management Activities - Storage Infrastructure Management Challenges.

Text Book:

1. EMC Corporation, Information Storage and Management, WileyIndia, 2nd Edition, 2011.

REFERENCES

2. Robert Spalding, "Storage Networks: The Complete Reference", Tata McGraw Hill, Osborne, 2003.
3. Marc Farley, Building Storage Networks, Tata McGraw Hill , Osborne,2nd Edition, 2001.
4. Meeta Gupta, Storage Area Network Fundamentals, Pearson Education Limited, 2002.

EIT721 ADVANCED DATA STRUCTURES LAB

L T P C

0 0 3 2

1. Write C++ programs to implement the following using an array.

a) Stack ADT b) Queue ADT.

2. Write C++ programs to implement the following using a singly linked list.

a) Stack ADT b) Queue ADT

3. Write C++ program to implement the deque (double ended queue) ADT using a doubly linked list.

4. Write a C++ program to perform the following operations:

a) Insert an element into a binary search tree.

b) Delete an element from a binary search tree.

c) Search for a key element in a binary search tree.

5. Write a C++ program to implement circular queue ADT using an array.

6. Write C++ programs that use non,recursive functions to traverse the given binary tree in

a) Preorder b) inorder and c) postorder.

7. Write a C++ programs for the implementation of bfs and dfs for a given graph.

8. Write C++ programs for implementing the following sorting methods:

a) Quick sort b) Merge sort c) Heap sort

9. Write a C++ program to perform the following operations

a) Insertion into a B,tree b) Deletion from a B,tree

10. Write a C++ program to perform the following operations

a) Insertion into an AVL,tree b) Deletion from an AVL,tree

11. Write a C++ program to implement Kruskal's algorithm to generate a minimum spanning tree.

12. Write a C++ program to implement Prim's algorithm to generate a minimum spanning tree.

13. Write a C++ program to implement all the functions of a dictionary (ADT) using hashing.

ECS 723 DATA ANALYTICS LAB

L T P C

0 0 3 2

1. Twitter Data download using python
2. YouTube Data download using python
3. Statistical analysis with twitter data
4. Sentiment analysis with twitter
5. Text mining with twitter
6. Training linear regression for prediction using UCI machine learning data set
7. Create CTR prediction pipeline using spark
8. Exploratory data analysis using PCA and feature based aggregation
9. Train logistic regression using stochastic gradient ascent

EIT706 NOSQL DATABASES

L T P C

4 0 0 4

Module I **11 Hours**

Big Data ,Scalability, Sorted Ordered Column-Oriented Stores: Key/Value Stores, Document Databases Graph Databases Examples, Storing and Accessing Data, Storing Data In and Accessing Data from MongoDB, Querying MongoDB, Storing Data In and Accessing Data from Redis Querying Redis, Storing Data In and Accessing Data from HBase, Querying HBase Storing Data In and Accessing Data from Apache Cassandra, Querying Apache Cassandra Language Bindings for NoSQL Data Stores, Language Bindings for Java Language Bindings for Python, Language Bindings for Ruby, Language Bindings for PHP

Module II **10 Hours**

Working with Column-Oriented Databases, Contrasting Column Databases with RDBMS, Column Databases as Nested Maps of Key/Value Pairs, Laying out the Web table, HBase Distributed Storage Architecture Document Store Internals , Storing Data in Memory-Mapped Files, Guidelines for Using Collections and Indexes in MongoDB, MongoDB Reliability and Durability ,Horizontal Scaling, Understanding Key/Value Stores in Memcached and Redis Under the Hood of Memcached Redis Internals , Eventually Consistent Non-relational Databases Consistent Hashing Object Versioning, Gossip-Based Membership and Hinted Handoff

Module III **10 Hours**

Creating Records, Creating Records in a Document-Centric Database, Using the Create Operation in Column-Oriented Databases, Using the Create Operation in Key/Value Maps

Accessing Data, Accessing Documents from MongoDB, Accessing Data from HBase, Querying Redis, Updating and Deleting Data , Updating and Modifying Data in MongoDB, HBase, and Redis, Limited Atomicity and Transactional Integrity

Module IV **10 Hours**

Similarities between SQL and MongoDB Query Features, Map Reduce in MongoDB, Accessing Data from Column-Oriented Databases like HBase, Querying Redis Data Stores, Changing Document Databases, Schema-less Flexibility, Exporting and Importing Data from and into MongoDB, Schema Evolution in Column-Oriented Databases, HBase Data Import and Export, Data Evolution in Key/Value Stores

Module V **10 Hours**

Essential Concepts Behind a Database Index, Indexing and Ordering in MongoDB, Creating and Using Indexes in MongoDB, Compound and Embedded Keys, Creating Unique and Sparse Indexes, Keyword-based Search and Multi Keys, Indexing and Ordering in CouchDB, The B-tree Index in CouchDB, Indexing in Apache Cassandra, Distributed ACID Systems, Consistency, Availability, Partition Tolerance, Upholding CAP, Compromising on Availability, Compromising on Partition Tolerance, Compromising on Consistency, Consistency Implementations in a Few NoSQL Products

Text Book:

1. Shashank Tiwari, Professional NoSQL, Wiley- August 2011

References

1. Dan MC Creary and Ann Kelly, Making Sense of NoSQL: A guide for Managers and the Rest of Us, Dreamtech Press, 2013

2. Adam Fowler, “Nosql for Dummies”, Wiley.

EIT708 BIG DATA AND ANALYTICS

L T P C

4 0 0 4

Module I **8 Hours**

Big Data's fundamental concepts, Understanding the business motivations and drivers behind Big Data adoption, characteristics of big datasets, big data technologies, applications of big data.

MapReduce, The Hadoop Distributed File system, Yarn, Hadoop I/O

Module II **10 hours**

MapReduce: Developing a MapReduce Application: configuration APIs, Setting Up the Development Environment, Writing a Unit Test with MRUnit, Running Locally on Test Data, Running on a Cluster- Packaging a Job, Launching a Job, The MapReduce Web UI, Retrieving the Results, Debugging a Job, Hadoop Logs, Remote Debugging; Tuning a Job, MapReduce Workflows- Decomposing a Problem into MapReduce Jobs, JobControl, Apache Oozie.

How MapReduce Works: Anatomy of a MapReduce Job Run, Failures, Shuffle and Sort, Task Execution

Module III **10 hours**

Pig: Pig Latin-Structure, Statements, Expressions, Types, Schemas, Functions, Macros, User-Defined Functions- A Filter UDF, An Eval UDF, A Load UDF; Data Processing Operators- Loading and Storing Data, Filtering Data, Grouping and Joining Data, Sorting Data, Combining and Splitting Data, Pig in Practice-- Parallelism, Anonymous Relations, Parameter Substitution

Module IV **9 hours**

Hive : HiveQL- Data Types, Operators and Functions, Tables-Managed Tables and External Tables, Partitions and Buckets, Storage Formats, Importing Data, Altering Tables, Dropping Tables, Querying Data- Sorting and Aggregating, MapReduce Scripts, Joins, Sub queries, Views User-Defined Functions.

Module V **10 hours**

Spark : Resilient Distributed Datasets- Creation, Transformations and Actions, Persistence, Serialization, Shared Variables, Anatomy of a Spark Job Run--Job Submission, DAG Construction, Task Scheduling, Task Execution; Executors and Cluster Managers, Further Reading

Text Book(s)

1. Tom White, Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale, 4/e, O'Reilly.

References

1. Seema Acharya, Subhashini Chellappan, Big Data and Analytics, Wiley India Pvt.Ltd

EIT710 SOCIAL NETWORK ANALYSIS

L T P C

4 0 0 4

Module I **9 hours**

Networks and Relations, Relations and Attributes, Analysis of Network Data, Interpretation of Network Data, An Overview. The Development of Social Network Analysis, Sociometric analysis and Graph Theory, Interpersonal Configurations and cliques, Towards formal models and structure.

Module II **9 hours**

Analyzing Relational Data, Collecting Relational Data, Selection and Sampling of Relational Data, Preparation of Relational Data, Organizing Relational Data. Lines, Neighborhoods and Densities, Sociometric and Graph Theory, Density: Ego-centric and Socio-centric, A Digression on absolute density, Community Structure and density.

Module III **9 hours**

Centrality Peripherality and Centralization, Centrality: Local and Global, Centralization and Graph Centres, bank Centrality in Corporate Networks, Components, Cores and Cliques, Components, Cycles and Knots, The Contours of components, Cliques and their intersections, Components and citation circles

Module IV **9 hours**

Positions, sets and clusters, the structural equivalence of points, Clusters: Combining and dividing points, Block Modeling with CONCER, Towards Regular Structure Equivalence, Corporate interlocks and Participations.

Module V **9 hours**

Network Dynamics and Change over Time, Modeling change in Network Structure, Testing Explanations. Dimensions and displays, Distance, space and metrics, principal components and factors, Non -metric methods, Advances in Network Visualization, Elites, Communities and influence. Accessing twitter API, Discovering the Trend Topics, LinkedIn

Text Book(s)

1. John Scott, Social Network Analysis, 3/e, SAGE Publications Ltd.

2. Matthew A. Russell , Mining the Social Web, 2nd Edition**,** O'Reilly Media

References

1. Charles Kadushin , Understanding Social Networks: Theories, Concepts, and Findings

2. Maksim Tsvetovat, Alexander Kouznetsov , Social Network Analysis for Startups, O'REILLY

EIT752 STREAM MINING

(Program Elective III)

L T P C

3 0 0 3

Module I **9 hours**

Introduction, Data Mining and Data Streams ,Data Stream Models, Basic Streaming Methods, Illustrative Applications

Module II **9 hours**

Introduction to Change Detection, Tracking Drifting Concepts, Monitoring the Learning Process Final Remarks, Maintaining Histograms from Data Streams**:** Introduction, Histograms from Data Streams, The Partition Incremental Discretization (PiD) Algorithm, Applications to Data Mining, Evaluating Streaming Algorithms **,** Learning from Data Streams, Evaluation Issues, Lessons Learned and Open Issues

Module III **9 hours**

Clustering from Data Streams, Clustering Examples, Clustering Variables, Frequent Pattern Mining, Frequent item set Mining, Heavy Hitters, Mining Frequent Item sets from Data Streams Sequence Pattern Mining, Decision Trees from Data Streams**,** The Very Fast Decision Tree Algorithm, Extensions to the Basic Algorithm, OLIN: Info-Fuzzy Algorithms, Novelty Detection in Data Streams, Learning and Novelty, Novelty Detection as a One-Class Classification Problem, Learning New Concepts, The Online Novelty and Drift Detection Algorithm

Module IV **9 hours**

Ensembles of Classifiers**,** Linear Combination of Ensembles, Sampling from a Training Set, Ensembles of Trees, Adapting to Drift Using Ensembles of Classifiers, Mining Skewed Data Streams with Ensembles, Time Series Data Streams**,** Time Series Analysis, Time Series Prediction, Similarity between Time Series, Symbolic Approximation (SAX)

Module VI **9 hours**

Ubiquitous Data Mining, Distributed Data Stream Monitoring, Distributed Clustering, Algorithm Granularity, the Next Generation of Knowledge Discovery

Textbook:

1. Joao Gama, Knowledge Discovery from Data Streams, A Chapman& Hall Book

2. Kapil Wankhade,Snehlata Dongre,Data Streams Mining

EIT742 BIOINFORMATICS

(Program Elective III)

L T P C

3 0 0 3

Module I **10 Hours**

Introduction: Basic biology, genetic material, genes, what molecules code for genes, structure of DNA, what carries information between DNA and proteins, proteins, analysis of DNA, why bioinformatics.

Module II **10 Hours**

Exhaustive Search: Restriction mapping, impractical restriction mapping algorithm , practical Restriction mapping algorithm, Regulatory motifs in DNA sequences, profiles, the motif finding problem, search trees, finding motifs , finding a median string.

Module III **8 Hours**

Greedy Algorithms: Genome rearrangement, sorting by reversals, approximation algorithm, breakpoints, greedy approach for motif finding.

Module IV **10 Hours**

Dynamic Programming Algorithm: Edit distance and assignments, longest common subsequence, global sequence alignment, scoring alignment, local sequence alignment, alignment with gap penalties, multiple alignment, gene prediction, statistical approach to gene prediction, Similarity based approach to gene prediction

Module V **10 Hours**

Clustering and Trees: Gene expression analysis, hierarchical clustering, K-mean clustering, clustering and corrupted cliques, evolutionary tree, distance based tree construction, reconstructing tree for additive matrices, evolutionary tree and hierarchical clustering, character based tree clustering.

Text Book(s)

1. Jones, N. and Pevzner, P., An introduction to Bioinformatics Algorithms, MITPress, 2004.

References

1. Durbin, R., Eddy, S., Krogh, A., Mitchison, G. , Biological Sequence Analysis: Probabilistic models of proteins and nucleic acids, Cambridge University Press,1998

2. Gusfield, D., Algorithms on Strings, Trees, and Sequences.Cambridge University Press,1997.

3. Waterman, M., Introduction to Computational Biology: Maps, sequences and genomes, CRC Press,1995.

4. Aluru, S.,Handbook of Computational Molecular Biology, CRC Press,2006.

EIT754 MINING MASSIVE DATASETS

(Program Elective III)

L T P C

3 0 0 3

Module I **8 hours**

Distributed Computing: Mining frequent patterns, associations, and correlations: basic concepts and methods, frequent item set mining methods, distributed computing, distributed data association rule mining, parallel and distributed data mining, supervised learning pipeline

Module II **8 hours**

Linear regression, distributed machine learning computation and storage, gradient descent, communication hierarchy, distributed machine learning communication principles

Module III **8 hours**

Linear classification and logistic regression probabilistic interpretation, using probabilistic predictions, categorical data and one hot encoding ,computing and storing OHE features, feature hashing

Module IV **8 hours**

PCA overview, PCA assumptions and solutions, PCA algorithm, PCA derivation, distributed PCA

Module V **8 hours**

Distributed data clustering, improved distributed combining algorithm, distributed clustering algorithm, distributed hierarchical clustering

Text Book(s)

1. Ron Bekkerman, Mikhail Bilenko, Scaling up Machine learning: Parallel and Distributed Approaches, Cambridge
2. Kimito Funatsu, New Fundamental Technologies in Data Mining, InTech

References:

1.Ali, Data mining methods and techniques, Cengage.

EIT756 INFORMATION RETRIEVAL
(Program Elective IV)

L T P C

3 0 0 3

Module I **9 hours**

Introduction -History of IR- Components of IR – Issues –Open source Search engine Frameworks – The impact of the web on IR – The role of artificial intelligence (AI) in IR – IR Versus Web Search – Components of a Search engine- Characterizing the web.

Module II **9 hours**

Information Retrieval: Boolean and vector-space retrieval models- Term weighting – TF-IDF weighting- cosine similarity – Preprocessing – Inverted indices – efficient processing with sparse vectors – Language Model based IR – Probabilistic IR –Latent Semantic Indexing – Relevance feedback and query expansion.

Module III **9 hours**

Web Search Engine – Introduction And Crawling Web search overview, web structure, the user, paid placement, search engine optimization/ spam. Web size measurement – search engine optimization/spam – Web Search Architectures – crawling – meta-crawlers- Focused Crawling – web indexes –- Near-duplicate detection – Index Compression – XML retrieval.

Module IV **9 hours**

Web Search – Link Analysis And Specialized Search: Link Analysis –hubs and authorities – Page Rank and HITS algorithms -Searching and Ranking – Relevance Scoring and ranking for Web – Similarity – Hadoop & Map Reduce – Evaluation – Personalized search – Collaborative filtering and content-based recommendation of documents and products – handling “invisible” Web – Snippet generation, Summarization, Question Answering, Cross- Lingual Retrieval.

Module V **9 hours**

Text Mining: Information filtering; organization and relevance feedback – Text Mining -Text classification and clustering – Categorization algorithms: naive Bayes; decision trees; and nearest neighbor – Clustering algorithms: agglomerative clustering; k-means; expectation maximization (EM).

Text Books:

1. C. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval , Cambridge University Press 2008.

2. Ricardo Baeza -Yates and Berthier Ribeiro – Neto, Modern Information Retrieval: The Concepts and Technology behind Search 2nd Edition, ACM Press Books 2011.

3. Bruce Croft, Donald Metzler and Trevor Strohman, Search Engines: Information Retrieval in Practice, 1st Edition Addison Wesley, 2009.

4. Mark Levene, An Introduction to Search Engines and Web Navigation, 2nd Edition Wiley, 2010.

References:

1. Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack, Information Retrieval: Implementing and Evaluating Search Engines, The MITPress ,2010.

2. Ophir Frieder “Information Retrieval: Algorithms and Heuristics: The Information Retrieval Series“, 2nd Edition,Springer,2004.

3. Manu Konchady, “Building Search Applications: Lucene, Ling Pipe”, and First Edition, Gate Mustru Publishing, 2008.

EIT758 DEEP LEARNING

(Program Elective IV)

L T P C

3 0 0 3

Module I **9 hours**

The Neural Network, Building Intelligent Machines, The Limits of Traditional Computer Programs, The Mechanics of Machine Learning, The Neuron, Expressing Linear Perceptrons as Neurons, Feed-forward Neural Networks, Linear Neurons and their Limitations, Sigmoid, Tanh, and ReLU Neurons, Softmax Output Layers, Looking Forward

Module II **9 hours**

Training Feed-Forward Neural Networks: The Cafeteria Problem, Gradient Descent, The Delta Rule and Learning Rates, Gradient Descent with Sigmoidal Neurons, The Backpropagation Algorithm, Stochastic and Mini-Batch Gradient Descent, Test Sets, Validation Sets, and Overfitting, Preventing Overfitting in Deep Neural Networks

Module III **10 hours**

Implementing Neural Networks in TensorFlow: What is TensorFlow? How Does TensorFlow Compare to Alternatives?, Installing TensorFlow, Creating and Manipulating TensorFlow Variables, TensorFlow Operations, Placeholder Tensors, Sessions in TensorFlow, Navigating Variable Scopes and Sharing Variables, Managing Models over the CPU and GPU, Specifying the Logistic Regression Model in TensorFlow, Logging and Training the Logistic Regression Model, Leveraging TensorBoard to Visualize Computation Graphs and Learning, Building a Multilayer Model for MNIST in TensorFlow

Module IV **10 hours**

Beyond Gradient Descent: The Challenges with Gradient Descent, Local Minima in the Error Surfaces of Deep Networks, Model Identifiability, How Pesky are Spurious Local Minima in Deep Networks?, Flat Regions in the Error Surface, When the Gradient Points in the Wrong Direction, Momentum-Based Optimization, A Brief View of Second Order Methods, Learning Rate Adaptation, AdaGrad - Accumulating Historical Gradients, RMSProp - Exponentially Weighted Moving Average of Gradients, Adam - Combining Momentum and RMSProp, The Philosophy Behind Optimizer Selection

Module V **10 hours**

Convolutional Neural Networks: Neurons in Human Vision, The Shortcomings of Feature Selection, Vanilla Deep Neural Networks Don't Scale, Filters and Feature Maps, Full Description of the Convolutional Layer, Max Pooling, Full Architectural Description of Convolution Networks, Closing the Loop on MNIST with Convolutional Networks, Image Preprocessing Pipelines Enable More Robust Models, Accelerating Training with Batch Normalization, Building a Convolutional Network for CIFAR-10, Visualizing Learning in Convolutional Networks, Leveraging Convolutional Filters to Replicate Artistic Styles, Learning Convolutional Filters for Other Problem Domains.

Text Book(s)

1. Nikhil Buduma, Fundamentals of Deep Learning, 1/e, O'relly

References

1. Julius Porter , Deep Learning: Fundamentals, Methods & Applications, Nova Science Publishers

EIT760 MULTIMEDIA DATA ANALYTICS

(Program Elective IV)

L T P C

3 0 0 3

Module I **10 Hours**

Multimedia Authoring and Data Representations:,multimedia Authoring and Tools, Graphics and Image Data Representations,Color in Image and Video,Fundamental Concepts in Video,Basics of Digital Audio

Module II **10 Hours**

II Multimedia Data Compression:Lossless Compression Algorithms,Lossy Compression Algorithms,Image Compression Standards,Basic Video Compression Techniques, MPEG Video Coding I - MPEG-1 and 2,MPEG Video Coding 11- MPEG-4, 7, and Beyond,Basic Audio Compression Techniques,MPEG Audio Compression

Module III **10 Hours**

Multimedia Communication and Retrieval:Computer and Multimedia Networks,Multimedia Network Communications and Application, Wireless Network, Content-Based retrieval in Digital Libraries

Module IV **10 Hours**

Multimedia Information sharing and and Retrieval:Social Media Sharing,cloud computing for Multimedia Services,Content-Based retrieval in Digital Libraries

Module V **10 Hours**

Machine Learning for Audio,Image and Video Analysis:From Perception to Computation.

Machine Learning:Taxonomy of Machine Learning ,Rote Learning , Learning from Instruction Learning by Analogy, Learning from Examples ,Supervised Learning ,Reinforcement Learning Unsupervised Learning

Text Books:

1.Fundamentals of Multimedia,By Ze-Nian Li, Mark S. Drew, Jiangchuan Li

2. Francesco Camastra,Alessandro Vinciarelli

EIT744 DIGITAL IMAGE PROCESSING

(Interdisciplinary Elective II)

L T P C

3 0 0 3

Module I **10 Hours**

Introduction: What is Digital Image Processing, Examples of fields that use digital image processing, fundamental steps in digital image processing, components of image processing system.. Digital Image Fundamentals: A simple image formation model, image sampling and quantization, basic relationships between pixels.

Module II **10 Hours**

Image Enhancement in the Spatial Domain: Basic gray, level transformation, histogram processing, enhancement using arithmetic and logic operators, basic spatial filtering, smoothing and sharpening spatial filters.

Module III **10 Hours**

Image Restoration: A model of the image degradation/restoration process, noise models, restoration in the presence of noise–only spatial filtering, Weiner filtering, constrained least squares filtering, geometric transforms; Introduction to the Fourier transform and the frequency domain, estimating the degradation function. Color Image Processing- Color fundamentals, color models.

Module IV **10 Hours**

Image Compression: Fundamentals, image compression models, Lossless Compression- Huffman coding, Run length coding contour coding,a brief discussion on Lossy Compression Image compression standards. Morphological Image Processing: Preliminaries, dilation, erosion, open and closing, hit or miss transformation, basic morphologic algorithms.

Module V **8 Hours**

Image Segmentation: Detection of discontinuous, edge linking and boundary detection, threshold, region–based segmentation.

Text Book(s)

1. Rafeal C. Gonzalez, Richard E. Woods, Digital Image Processing, 2/e, Pearson Education, 2002.

References

1. Milan Sonka, Vaclav Hlavac and Roger Boyle, Image Processing, Analysis, and Machine Vision, 2/e, Thomson Learning, 2007.
2. Alasdair Mc Andrew , Introduction to Digital Image Processing with Matlab- Thomson Course Technology, Brooks/Cole,2004.
3. Adrian Low , Computer Vision and Image Processing, , 2/e, B.S.
4. Publications,2008.
5. Rafeal C.Gonzalez, Richard E.Woods, Steven L. Eddins, Digital Image
6. Processing using Matlab, Pearson Education,2003.

ECS749 INTERNET OF THINGS

(Interdisciplinary Elective II)

L T P C

3 0 0 3

Module I **8 hours**

Introduction: The Internet of Things: An overview, the flavour of the internet of things, the "internet" of "things", the technology of the internet of things, enchanted objects, who is making the internet of things. **Design Principles for Connected Devices:** Calm and ambient technology, magicas metaphor, privacy, web thinking for connected devices, affordances.

Module II **10 hours**

Internet Principles: Internet communications: An overview (IP, TCP, the IP protocol suite (TCP/IP), UDP), IP addresses (DNS, Static IP Address assignment, dynamic IP address assignment, IPv6), MAC addresses, TCP and UDP ports, application layer protocols.

Module III **10 hours**

Prototyping : Thinking about Prototyping: Sketching, familiarity, costs versus ease of prototyping, prototypes and production, open source versus closed source, tapping into the community. **Prototyping Embedded Devices:** Electronics, embedded computing basics, developing on the arduino, raspberry pi, beaglebone black, electric imp, mobile phone and tablets, plug computing, always on internet of things.

Module IV **10 hours**

Prototyping the Physical Design: Preparation, sketch, iterate and explore, non digital methods, laser cutting, 3D printing, CNC milling, repurposing/ recycling. **Techniques for Writing Embedded Code:** Memory Management, performance and battery life, libraries, debugging.

Module V **10 hours**

Prototype to Reality: Business Models A short history of business models, the business model canvas, models, funding an internet of things startup, lean startups. **Moving to manufacture:** Designing kits, designing printed circuit boards, manufacturing printed circuit boards, mass, producing the case and other fixtures, certification, costs, scaling up software.

Text Book(s)

1. Adrian McEwen, Hakim Cassimally, Designing the Internet of Things, Wiley, 2013.

References

1. Charalampos Doukas, Building Internet of Things with the Arduino, Create space, 2002.
2. Dieter Uckelmann et.al, Architecting the Internet of Things, Springer, 2011.
3. Luigi Atzor et.al, The Internet of Things: A survey, Journal on Networks, Elsevier, 2010.

EID777 GRAPH THEORY & ALGORITHMS

(Interdisciplinary Elective II)

L T P C

3 0 0 3

Module I **10 hours**

Review of basic notions in graph theory, algorithms and complexity: Graphs – Introduction, Basic graph theoretic definitions– Isomorphism – Sub graphs – Walks, Paths, Circuits – Connectedness – Components – Euler graphs – Hamiltonian paths and circuits Graph representations. Classes P and NP, NP-hardness, polynomial reductions, 2-SAT problem, 3-SAT problem.

Module II **10 hours**

Trees, Connectivity & Planarity: Spanning trees – Fundamental circuits – Spanning trees in a weighted graph – Minimal spanning tree algorithm - Kruskal and Prim's algorithm - Shortest path algorithms – Dijsktra's algorithm - DFS and BFS algorithms.

Cut sets – Properties of cut set – All cut sets – Fundamental circuits and cut sets – Connectivity and separability – Network flows – 1-Isomorphism – 2-Isomorphism – Combinational and geometric graphs – Planer graphs – Different representation of a planer graph.

Module III **10 hours**

Graph colorings: Chromatic number, upper and lower bounds. Greedy algorithm and its analysis. The Four Color Theorem. Hadwiger's Conjecture. Brooks' Theorem. Edge colorings and Vizing's Theorem. List colorings. Galvin's Theorem. Algorithmic aspects of graph coloring. NP-completeness of the problem of computing the chromatic number. Applications in scheduling. **Approximation algorithms for graph problems:** Approximation algorithm for vertex cover problem. Approximation algorithms for the metric traveling salesman problem.

Module IV **10 hours**

Perfect graphs and their subclasses: Basic theory and examples of hereditary graph classes. Perfect graphs and their properties. Cographs. Split graphs and threshold graphs. Chordal graphs. Interval graphs. Efficient algorithms for various problems based on structural properties of graphs in these classes.

Module V **10 hours**

Further examples of tractable problems: Polynomial time algorithm for the maximum cut problem in planar graphs. Polynomial time algorithm for the 3-coloring problem on graphs with small dominating sets. The independent set problem: Matching techniques. Method of augmenting graphs. Decomposition by clique separators. Modular decomposition. Bounded tree-width, bounded clique-width. Applications of these methods, both individually and combined.

Text Books

1. Narsingh Deo, "Graph Theory: With Application to Engineering and Computer Science", Prentice Hall of India, 2003.

2. Clark J. and Holton D.A, "A First Look at Graph Theory", Allied Publishers, 1995.

1. Rosen K.H., "Discrete Mathematics and Its Applications", McGraw Hill, 2007.

2. Mott J.L., Kandel A. and Baker T.P. "Discrete Mathematics for Computer Scientists and Mathematicians" , Prentice Hall of India, 1996.

References:

1.T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to Algorithms, second edition McGraw-Hill, 2001.

2. M. R. Garey, D. S. Johnson, Computers and Intractability: A guide to the theory of NPCompleteness,1979.

3. M. C. Golumbic. Algorithmic Graph Theory and Perfect Graphs, Volume 57 in the series Annals of Discrete Mathematics. North Holland, second edition, 2004.

4. A. Brandstädt, V. B. Le, J. P. Spinrad. Graph Classes: A Survey. SIAM, 1999.

5. V. V. Vazirani. Approximation Algorithms, Springer-Verlag, 2001.

6. J. A. Bondy, U. S. R. Murty. Graph Theory, North-Holland, Springer-Verlag, 2008.

7. B. Korte, J. Vygen. Combinatorial Optimization. Theory and Algorithms, Volume 21 in the series Algorithms and Combinatorics, Springer-Verlag, druga izdaja, 2002.

8. R. Diestel, Graph Theory. Springer, 2006.

9. A. Schrijver, Combinatorial Optimization, Springer, 2003.

EIT726 HADOOP ECOSYSTEMS LAB

L	T	P	C
0	0	3	2

1. Map reduce program for word count

2. PIG

1. Working with Pig commands
2. Pig execution modes, Diagnostic operators-dump, describe, explain, illustrate
3. Load and store data from various file formats
4. Grouping and joins-join, co group ,group, cross
5. Filter, distinct, for each, generate
6. Sort-order ,limit, Union, split
7. Write Pig script for word count

3. Hive

1. Working on Hive commands
2. Creation of manged, external tables
3. Load data form files, load data from another table, load data from another table during table creation
4. Sub queries,joins,group by ,aggregate operators
5. Static, dynamic partitioning
6. Import tables from MySQL to Hive using Sqoop
7. Convert unstructured data to structured data using SERDE, regular expressions

Spark

1. Working with Spark commands like map, reduce, filter, groupBy, sort etc
2. Function to find sum of each column of given set
3. Running Clustering algorithms in Spark
4. Running Classification algorithms in Spark

EIT728 BIG DATA ANALYTICS LAB

L T P C

0 0 3 2

1. Compute simple linear regression weights and make predictions of the output Given the input feature
2. Compute multiple regression weights and predict the output given the input feature and also compute error
3. Compute the regression weights using gradient descent algorithm given initial weight vector, step size ,tolerance
4. Implement ridge regression to compute the polynomial regression and find Best L2 penalty using
   cross validation
5. Implement Gradient Descent algorithm for ridge regression given initial weight vector, step size, tolerance
6. Run the Lasso regression with different L1 penalties and choose best L1 penalty using validation set
7. Implement coordinate descent algorithm for LASSO
8. Implement the gradient ascent for Logistic Regression
9. Write functions to transform categorical features into binary features and build binary decision tree and evaluate the accuracy
10. Implement binary decision trees with different early stopping methods