REGULATIONS & SYLLABUS

Master of Technology
in
Data Science
(W.e.f 2017-18 admitted batch)

Website: www.gitam.edu
### Semester –I

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

### Semester –II

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

### Semester –III

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EIT891</td>
<td>Project Work-I</td>
<td>PW</td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>EIT893</td>
<td>Comprehensive Viva-voce</td>
<td>CE</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

### Semester- IV

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EIT892</td>
<td>Project work-II</td>
<td>PW</td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
</tbody>
</table>

2
### Interdisciplinary Elective-I

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EID774</td>
<td>Fraud analytics</td>
<td>IDE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>EID772</td>
<td>E-commerce</td>
<td>IDE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>EIT741</td>
<td>Information Storage management</td>
<td>IDE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Interdisciplinary Elective-II

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EIT744</td>
<td>Digital Image Processing</td>
<td>IDE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>ECS749</td>
<td>Internet of Things</td>
<td>IDE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>EID777</td>
<td>Graph theory and Algorithms</td>
<td>IDE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Program Elective-I

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EIT749</td>
<td>Optimization Techniques</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>EIT751</td>
<td>Exploratory Data analysis</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>EIT753</td>
<td>Game Theory</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Program Elective-II

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ECS746</td>
<td>Natural Language Processing</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>EIT755</td>
<td>Web Semantics</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>ECS748</td>
<td>Cloud computing</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Program Elective-III

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EIT752</td>
<td>Stream Mining</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>EIT742</td>
<td>Bioinformatics</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>EIT754</td>
<td>Mining Massive Data Sets</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Program Elective-IV

<table>
<thead>
<tr>
<th>S.No</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Category</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EIT756</td>
<td>Information Retrieval</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>EIT758</td>
<td>Deep Learning</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>EIT760</td>
<td>Multimedia Data analytics</td>
<td>PE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>
### Number of Credits

<table>
<thead>
<tr>
<th>Semester</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credits</td>
<td>25</td>
<td>27</td>
<td>10</td>
<td>14</td>
<td>76</td>
</tr>
</tbody>
</table>
Module I


Module II


Module III


Module IV

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


Text Book:
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
Module I  12hrs


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


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


Text Book(s)


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

References


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

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

Module III

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

Module V

Text Book(s)


References


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

Module II
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

Module IV
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
Module I 10 hours

Module II 10 hours

Module III 10 hours

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
Module I


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


Module III

**Probabilistic Context Free Grammars**


**Module V**

10 hours

**Maximum Entropy Markov Models & Conditional Random Fields**


**Text Books:**

1. "Speech and Language Processing": Jurafsky and Martin, Prentice Hall
2. "Statistical Natural Language Processing": Manning and Schutze, MIT Press

**References:**

4. Lutz and Ascher - "Learning Python", O'Reilly
Module I
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

Module III
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
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
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)
References

ECS748 CLOUD COMPUTING
(Program Elective II)

L T P C
3 0 0 3

Module I

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

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

Module III

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

Module IV

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

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)

References


## EID774 FRAUD ANALYTICS
(Interdisciplinary Elective I)

<table>
<thead>
<tr>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

### Module I
**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
**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
**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
**Predictive Analytics for Fraud Detection**

### Module V
**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:**

EID 772 E-COMMERCE
(Interdisciplinary Elective I)

<table>
<thead>
<tr>
<th>Modules</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module I</td>
<td>10</td>
</tr>
<tr>
<td>Module II</td>
<td>10</td>
</tr>
<tr>
<td>Module III</td>
<td>10</td>
</tr>
<tr>
<td>Module IV</td>
<td>10</td>
</tr>
<tr>
<td>Module V</td>
<td>10</td>
</tr>
</tbody>
</table>

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)

References
2. Efrain Turbon, Jae Lee, David King, H. Michael Chang, E-Commerce, 3/e, Pearson
EIT741 INFORMATION STORAGE MANAGEMENT
(Interdisciplinary Elective I)

L T P C
3 0 0 3

MODULE I
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
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
BACKUP AND RECOVERY
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
CLOUD COMPUTING
Computing - Cloud Service Models-Cloud Deployment models-Cloud computing
Infrastructure-Cloud Challenges.
MODULE V

SECURING AND MANAGING STORAGE INFRASTRUCTURE


Text Book:


REFERENCES


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.
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

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

Text Book:
1. Shashank Tiwari, Professional NoSQL, Wiley- August 2011

References
Module I
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
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
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
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
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.
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)

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

**Module II**

**Module III**
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**
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**
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)

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

Module IV
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
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)

References


EIT754 MINING MASSIVE DATASETS
(Program Elective III)

<table>
<thead>
<tr>
<th>Module</th>
<th>L T P C</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module I</td>
<td>3 0 0 3</td>
<td>8</td>
</tr>
</tbody>
</table>

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.
Module I
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
Information Retrieval: Boolean and vector-space retrieval models - Term weighting – TF-IDF
weighting- cosine similarity – Preprocessing – Inverted indices – efficient processing with sparse
feedback and query expansion.

Module III
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 –

Module IV
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
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:**


**References:**

EIT758 DEEP LEARNING
(Program Elective IV)

L T P C
3 0 0 3

Module I


Module II

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


Module IV

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

EIT760 MULTIMEDIA DATA ANALYTICS
(Program Elective IV)

Module I
10 Hours

Module II
10 Hours
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 II - 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 Retrieval: Social Media Sharing, cloud computing for Multimedia Services, Content-Based retrieval in Digital Libraries

Module V
10 Hours

Text Books:
1. Fundamentals of Multimedia, By Ze-Nian Li, Mark S. Drew, Jiangchuan Li
2. Francesco Camastra, Alessandro Vinciarelli
Module I


Module II

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

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

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

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

Text Book(s)

References


3. Adrian Low, Computer Vision and Image Processing, 2/e, B.S.


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)  

References  
1. Charalampos Doukas, Building Internet of Things with the Arduino, Create space, 2002.  

46
EID777 GRAPH THEORY & ALGORITHMS
(Interdisciplinary Elective II)

<table>
<thead>
<tr>
<th>L</th>
<th>T</th>
<th>P</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

**Module I**

**10 hours**


**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**


**Module IV**

**10 hours**

**Perfect graphs and their subclasses:** Basic theory and examples of hereditary graph classes.

Module V

10 hours


Text Books


References:

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 managed, external tables
   3. Load data from 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
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