JSS MAHAVIDYAPEETHA
JSS SCIENCE AND TECHNOLOGY UNIVERSITY

Mysuru-570006.

Department of Information Science & Engineering

Master of Technology
In
Data Science

SCHEME & SYLLABUS

I to IV semesters

2019

Scheme of Teaching and Examination

MTech in Data Science

First Semester MTech(DS)

SL. No.	Subject Code	Course Title	Teaching Department	Credits				Contact Hours	Marks		
				L	T	P	Total		CIE	SEE	Total
1.	SDS110	Principles of Data Science	IS&E	4	1	0	5.0	6	50	50	100
2.	SDS120	Big Data Analytics	IS&E	4	1	0	5.0	6	50	50	100
3.	SDS130	Machine Learning	IS&E	4	0	1	5.0	6	50	50	100
4a.	SDS141	Computational Statistical Methods	IS&E	4	0	1	5.0	6	50	50	100
4b.	SDS142	Information Retrieval	IS&E	4	0	1	5.0	6	50	50	100
4c.	SDS143	Image & Video Analytics	IS&E	4	0	1	5.0	6	50	50	100
5a.	SDS151	Linear Algebra and Applications	IS&E	4	0	1	5.0	6	50	50	100
5b.	SDS152	System Security	IS&E	4	0	1	5.0	6	50	50	100
5c.	SDS153	Exploratory Data Analysis & Visualization	IS&E	4	0	1	5.0	6	50	50	100
6.	SDS160	Minor Project – 1	IS&E	0	0	3.0	3.0	6	100	-	100
Total	20	2.0	6.0	28	36	350	250	600			
		Total		20	2.0	6.0	28	36	350	250	600

Scheme of Teaching and Examination

MTech in Data Science

Second Semester MTech (DS)

SL. No.	Subject Code	Course Title	Teaching Department	Credits				Contact Hours	Marks		
				L	T	P	Total		CIE	SEE	Total
1.	SDS210	Advanced Data Mining Techniques	IS&E	4	1	0	5.0	6	50	50	100
2.	SDS220	Cloud Computing and Virtualization	IS&E	4	0	1	5.0	6	50	50	100
3.	SDS230	Deep Learning	IS&E	4	0	1	5.0	6	50	50	100
4a.	SDS241	Optimization Theory	IS&E	4	1	0	5.0	6	50	50	100
4b.	SDS242	Computational Linguistics	IS&E	4	1	0	5.0	6	50	50	100
4c.	SDS243	Bioinformatics	IS&E	4	1	0	5.0	6	50	50	100
5a.	SDS251	Scalable systems for Data Science	IS&E	4	0	1	5.0	6	50	50	100
5b.	SDS252	Web Databases & Information Systems	IS&E	4	0	1	5.0	6	50	50	100
5c.	SDS253	Social & Information Network Analysis	IS&E	4	0	1	5.0	6	50	50	100
6.	SDS260	Minor Project – 2	IS&E	0	0	3.0	3.0	6	100	-	100
Total	20	2.0	6.0	28	36	350	250	600			
		Total		20	2.0	6.0	28	36	350	250	600

Scheme of Teaching and Examination

MTech in Data Science

Third Semester MTech (DS)

Sl.No.	Subject Code	Course title	Teaching Department	Credits				Contact Hours	Marks			Exam Duration
				L	T	P	Total		CIE	SEE	Total	
1	SDS31T	Practical Training in Industry/Exploration in Research	IS&E	--	--	4	4	-	100	-	100	-
2	SDS32P	Project Work (Phase – I)	IS&E	--	--	10	10	-	100	-	100	-
				Total Credits			**14**		**Total Marks**		**200**	

Scheme of Teaching and Examination

MTech in Data Science

Fourth Semester MTech (DS)

Sl.No.	Subject Code	Course title	Teaching Department	Credits				Contact Hours	Marks			Exam Duration
				L	T	P	Total		CIE	SEE	Total	
1	SDS41P	Project Work (Phase –II)	IS&E	--	--	18	18	-	100	200	300	-
				Total Credits			**18**		**Total Marks**		**300**	

SDS110 PRINCIPLES OF DATA SCIENCE

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

On successful completion of the course students will have the ability to:

CO1: Explain a flow process for data science problems.

CO2: Obtain data to classify data science problems into standard typology using mathematical concepts.

CO3: Analyze results and Correlate them to the solution approach followed.

CO4: Design use cases to validate approach and identify modifications required and develop applications for data science solutions.

CO5: Assess the solution approach through visualization and presentation.

Introduction **10 Hours**

What is Data Science?, Basic Terminology, Why Data Science?, Example – Sigma Technologies, The data science Venn diagram, The math: Example – Spawner-Recruit Models, Computer programming (Preferably Python/R/Matlab/PERL), Some more terminology, Some Data science case studies, Data Models and its types. Types of Data, Flavors of data, Structured versus unstructured data, Quantitative versus qualitative data, The four levels of data, The Five Steps of Data Science.

Basic Mathematics – Vectors and Matrices **10 Hours**

Vectors and Linear Combinations, Lengths and Dot Products, Matrices, Solving Linear Equations: 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, Independence, Basis and Dimension, Dimensions of the Four Subspaces.

Probability **10 Hours**

Basic definitions, Probability, Bayesian versus Frequentist, Frequentist approach, Compound events, Conditional probability, The rules of probability, Collectively exhaustive events, Bayesian ideas revisited, Bayes theorem, Random variables.

Statistics **10 Hours**

Basic Statistics, What are statistics ?, How do we obtain and sample data?, Obtaining data, Sampling data, How do we measure statistics?, Point estimates, Sampling distributions, Confidence intervals, Hypothesis tests, Conducting a hypothesis test, Type I and type II errors, Hypothesis test for categorical variables.

Visualization **10 Hours**

Basic principles, ideas and tools for data visualization, why does communication matter?, Identifying effective and ineffective visualizations, Scatter plots, Line graphs, Bar charts, Histograms, Box plots, When graphs and statistics lie, Correlation versus causation, Simpson's paradox, If correlation doesn't imply causation, then what does?, Verbal communication, The why/how/what strategy of presenting, Data Science ethical issues.

Text Book:

1. Principles of Data Science, SinanOzdemir, PACKT Publisher, First Edition, 2016.

Reference Books:

1. Introduction to Linear Algebra, Gilbert Strang, Wellesley-Cambridge Press, Fifth Edition, 2016.
2. Doing Data Science: Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O'Reilly Media, 2013.
3. Mining of Massive Datasets, Jure Leskovec, AnandRajaraman, Jeff Ullman, Second Edition, Cambridge University Press Publisher, 2015.

Online Courses:

1. Swayam : https://swayam.gov.in/nd1_noc19_cs60/preview
2. NPTEL : https://nptel.ac.in/courses/106106179/

SDS120 BIG DATA ANALYTICS

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes (CO):

At the end of the course student will be able to:
CO1: Explain various forms of data and analytical models.
CO2: Apply big data technologies and tools to analyze the data.
CO3: Explain the various streaming techniques and recommendation system.
CO4: Apply machine learning algorithms to handle big data.
CO5: Understand the role of big data in the current industrial context.

Introduction **10 Hours**

Overview of Big Data, History, Structuring Big Data, Types of Data, Elements of Big Data, Data analytics project life cycle, Problems & challenges in understanding Data Analytics, Web page categorization, computing the frequency of stock market change. Use of Big Data in Social Networking, Use of Big Data in preventing Fraudulent activities, Use of Big Data in Retail Industry.

Big Data Technology **10 Hours**

Exploring Big Data Stack, Virtualization, Virtualization Approaches, Distributed and parallel computing for Big Data, The cloud and Big Data, Cloud Deployment Models, Cloud Delivery Models, Cloud providers in Big Data Market. Introducing Hadoop, Hadoop Ecosystem, Hadoop Distributed File Systems(HDFS), Features of HDFS : Hadoop YARN, MAP Reduce, Features of Map Reduce, Working of Map Reduce, Techniques to Optimize Map Reduce Jobs, Uses of Map Reduce, HBase, Features of HBase, Role of HBase in Big Data processing, Other tools of Hadoop (Hive, Pig and Pig Latin, Sqoop, ZooKeeper, Flume, OOZie),

Mining Data Streams **08 Hours**

The Stream Data Model, A Data-Stream-Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing, Sampling Data in a Stream, Filtering Streams, Estimating Moments, Dealing With Infinite Streams, Counting Ones in a Window.

Frequent Itemsets and Recommendation Systems **12 Hours**

The Market Basket Analysis, A Priori Algorithm, Handling Larger Datasets in Main Memory, Limited-Pass Algorithms, Counting Frequent Items in a Stream. Recommendation System: A Model, Content Based Recommendations, Collaborative Filtering, Dimensionality Reduction Problem, TheNetFlix Problem.

Large Scale Machine Learning **10 Hours**

Introduction, Types of Machine Learning Algorithms, Machine Learning Architecture, Applications of Machine Learning, Supervised Machine Learning Algorithms: Learning from Nearest Neighbors, Support Vector Machines. Unsupervised Machine Learning Algorithms: Hierarchical Clustering Techniques, K-means Algorithms.

Text Books:

1. Big Data: Black Book, DT Editorial Services, Dream Tech Press Publishers, 2015.
2. Mining of Massive Datasets, Jure Leskovec, AnandRajaraman, Jeff Ullman, Second Edition, Cambridge University Press Publisher, 2015.

Reference Books:

1. Big Data Analytics with R and Hadoop,VigneshPrajapati,Packt Publishing, 2013
2. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services, 2015.

SDS130 MACHINE LEARNING

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

On successful completion of the course students should be able to:
CO1: Understand the basic principles of Learning theories
CO2: Understand the principles of dimensionality reduction and feature selection techniques
CO3: Understand and Develop a wide variety of supervised learning algorithms
CO4: To become familiar with various clustering algorithms
CO5: To learn methodology and tools to apply machine learning algorithms to Software Engineering

Introduction & Bayesian Decision Theory **10 Hours**

What Is Machine Learning?, Challenges, Examples of Machine Learning Applications, Present Research Avenues, Introduction to Bayesian Decision Theory, Classification, Losses and Risks, Discriminant Functions, Utility Theory, Association Rules.

Dimensionality Reduction **10 Hours**

Introduction, Feature Generation, Feature Selection, Principal Component Analysis, Factor Analysis, Multidimensional Scaling, Linear Discriminant Analysis, Locality Preserving Projections (LPP) and its variants, Locality Preserving Indexing and its variants.

Supervised Learning **10 Hours**

Learning a Class from Examples, Probably Approximately Correct (PAC) Learning, Noise, Learning Multiple Classes, Regression, Model Selection and Generalization, Dimensions of a Supervised Machine Learning Algorithms, Decision Tree Induction, Nearest Neighbors, Bayesian Classifier, Model Over fitting, Performance Evaluation of classifiers.

Clustering **10 Hours**

Basic Concepts, Proximity Measures, Sequential Algorithms, Hierarchical Algorithms, Schemes based on Functional Optimization, Clustering Algorithms based on Graph Theory, Cluster Validity.

Multilayer Perceptron **10 Hours**

The Perceptron, Learning Boolean Functions, MLP as a universal approximator, Back Propagation Algorithm, Training Procedures, Tuning Networks, Recurrent Networks, Radial Basis functions.

Text Books:

1. **Introduction to Machine Learning,*EthemAlpaydin,*** Second Edition, PHI Learning Publisher, 2013 edition.
2. **Pattern Recognition,*SergiosTheodoridis and KonstantinosKoutroumbas,*** Fourth Edition, Academic Press Publisher, 2014.

Reference Materials:

1. **Machine Learning,*Tom M. Mitchell,***McGrawHil Publishers, 1997.
2. **Machine Learning Applications in Software Engineering, *Du Zhang and Jeffrey J. P. Tsai,*** World Scientific Publishers, 2005.
3. **Pattern Recognition and Machine Learning, *Christopher M. Bishop,***Spriger Publishers, 2011.
4. **Related Research Articles**

SDS141 COMPUTATIONAL STATISTICAL METHODS

Total Teaching Hours: 50 **No. of Credits : 05**

Course outcomes:

At the end of the course, students will have learned about the following concepts.
CO1: Explain computational statistical methods.
CO2: Illustrate unsupervised and supervised learning techniques.
CO3: Implement different learning algorithms.

Statistical Learning **10 hours**

Introduction, what is Statistical Learning, Assessing Model Accuracy

Linear Regression **10 hours**

Simple Linear Regression, Multiple Linear Regression, Other considerations in the Regression Model, The Market Plan, Comparison of Linear Regression with K-Nearest Neighbors

Classification **10 hours**

Overview of Classification, Why not Linear Regression, Logistics Regression, LDA, Comparison of Classification Methods

Resampling Methods/ Linear Model Selection & Regularization **10 hours**

Cross Validation, The Bootstrap Subset Selection, Shrinkage Methods, Dimensionality Reduction Methods, Considerations in High Dimension

Tree based Methods/SVM/Unsupervised Learning **10 hours**

Basics of Decision Trees, Bagging, Random Forests, Boosting SVMs: Maximal Margin Classifier, Support Vector Classifiers, SVMs, SVMs with more than two classes, Relation to Logistic Regression Unsupervised Learning: Challenges, PCA, Cluster Methods

Text Books:

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning with Applications in R”, 2014, Springer.
2. Geof Givens, Jennifer Hoeting, “Computational Statistics”, 2nd edition, 2013, Wiley

Reference Books:

1. Max Kuhn, Kjell Johnson, “Applied Predictive Modeling”, 2013, Springer

SDS142 INFORMATION RETRIEVAL

Total Teaching hours: 50 **No. of credits: 05**

Course outcomes:

On successful completion of the course, students should be able to:
CO1: Understand the basics of Information Retrieval (IR) models
CO2: Able to understand Boolean and vector-space retrieval models
CO3: Understand the importance of efficient Indexing techniques
CO4: Will be able to understand various searching techniques
CO5: Able to understand various feedback methods and evaluation metrics

Introduction:
10Hrs

Motivation, Basic concepts, Past, present, and future, The retrieval process. **Modeling:** Introduction, A taxonomy of information retrieval models, Retrieval: Adhoc and filtering, A formal characterization of IR models, Classic information retrieval, Alternative set theoretic models, Alternative algebraic models, Alternative probabilistic models, Structured text retrieval models, Models for browsing.

IR Using Boolean Model: **10 Hrs**

Information retrieval using the Boolean model, An example information retrieval problem, A first take at building an inverted index, Processing Boolean queries, The dictionary and postings lists, Determining dictionary terms, Postings lists, revisited, Tolerant retrieval, Wildcard queries, Spelling correction, Phonetic correction.

Indexing **10 Hrs**

Introduction; Inverted Files; Other indices for text; Boolean queries; Sequential searching; Pattern matching; Structural queries; Compression. Index construction, Construction of large indexes, Distributed indexing, Dynamic indexing, Other types of indexes
Parallel and Distributed IR: Introduction, Parallel IR, Distributed IR.

Vector space retrieval **10 Hrs**
Scoring and term weighting, Parametric and zone indexes, Weighted zone scoring, Term frequency and weighting, Inverse document frequency, tf-idf weighting, Variants in weighting functions, Documents as vectors, Inner products ,Queries as vectors, Heuristics for efficient scoring and ranking, Inexact top K document retrieval, Interaction between vector space and other retrieval methods , Query parsing and composite scoring, Text Operations: Introduction, Document preprocessing, Document classification and clustering.

Retrieval Evaluation **10Hrs**

Introduction, Retrieval performance evaluation, Reference collections. Query Languages: Introduction, keyword-based querying, Pattern matching, Structural queries, Query protocols. Query Operations: Introduction, User relevance feedback, Automatic local analysis, Automatic global analysis. Searching the Web: Introduction, Challenges, Characterizing the web, Search engines

Text Books:

1.Christopher D. Manning, PrabhakarRaghavan, HinrichSchütze: *Introduction to Information Retrieval,* Cambridge University Press, 2012

Reference Books:

1. Ricardo Baeza-Yates, BerthierRibeiro-Neto: *Modern Information Retrieval,* Pearson, 2003.
2. David A. Grossman, Ophir Frieder: *Information Retrieval Algorithms and Heuristics*, 2nd Edition, Springer, 2004.

SDS143 IMAGE AND VIDEO ANALYTICS

Total Teaching hours: 50 **No. of credits: 05**

Course outcomes:

On successful completion of the course, students should be able to:
CO1: Explain the fundamental principles of image and video analysis
CO2: Design the various filters for Image enhancement.
CO3: Apply Motion Estimation techniques for Video
CO4: Apply various compression methods for video and Image
CO5: Analyze various segmentation and Compression techniques

Introduction **10 Hours**

Digital Image Fundamentals, Elements of visual perception, Image sensing and acquisition, Sampling and Quantization, Relationships between pixels, Linear and Non-Linear operations.
Image and Video Processing: Basic Linear Filtering to Image Enhancement, Non-Linear Filtering for Image Analysis and Enhancement, Morphological Filtering for Image Enhancement and Detection, Image Restoration, Motion Detection and Estimation, Video enhancement and restoration.

Image Enhancement **10 Hours**

Basic Gray Level Transformations, Histogram Processing, Enhancement using Arithmetic and Logical Operations, Basics of Spatial Filtering, Image Enhancement in Frequency Domain, Smoothing and Sharpening Frequency Domain Filters, Homomorphic Filters.

Digital Video & Motion Estimation **10 Hours**

Human Visual System and Color, Analog and Digital Video, 3D Video, Digital Video Applications, Image and Video quality, Motion Models, 2D Apparent Motion estimation, Differential Methods, Matching Methods, Non-Linear Optimization Methods, Transform Domain Methods.

Image and Video Segmentation **10 Hours**

Detection of Discontinues, Edge Linking and Boundary Detection, Threshold Based Segmentation, Region Based Segmentation, Segmentation by Morphological Watersheds, The use of Motions in Segmentation, Change Detection, Motion Tracking, Image and Video Matting.

Image & Video Compression **10 Hours**

Basics of Image Compression, Lossless Image Compression, Discrete Cosine Transform Coding and JPEG, Wavelet Transform Coding and JPEG2000, Video Compression Approaches, Early Video Compression Standards, MPEG-4, H.264 Standard, High Efficiency Video Coding (HEVC) Standard, Scalable Video

Compression.

Text Books:

1.Digital Video Processing, ***Murat Tekalp,*** Second Edition, Prentice Hall, 2015.

2.Digital Image Processing, ***Rafael C. Gonzalez and Richard E. Woods,*** Fourth Edition, Pearson Publisher, 2017.

Reference Book:

1. Handbook of Image and Video Processing, ***Alan C. Bovik,*** Second Edition, Academic Press, 2005.

SDS151 LINEAR ALGEBRA AND APPLICATIONS

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

On successful completion of the course, students should be able to:
CO1: Solve linear equations through matrix representation.
CO2: Identify the vector space and subspace.
CO3: Test for orthogonality and apply Eigen vectors to solve differential equations.
CO4: Justify linear inequalities in the vector subspace.
CO5: Apply methods to solve different linear equations.

Introduction **10 Hours**

Introduction, The Geometry of Linear Equations, An Example of Gaussian Elimination, Matrix Notation and Matrix Multiplication, Triangular Factors and Row Exchanges, Inverses and Transposes, Special Matrices and Applications.

Vector Space **10 Hours**

Vector Spaces and Subspaces, Solving Ax= 0 and $Ax = b$, 3 Linear Independence, Basis, and Dimension, The Four Fundamental Subspaces, Graphs and Networks, Linear Transformations, Review Exercises.

Orthogonality **10 Hours**

Orthogonal Vectors and Subspaces, Cosines and Projections onto Lines, Projections and Least Squares, Orthogonal Bases and Gram-Schmidt, The Fast Fourier Transform.

Computations with Matrices **10 Hours**

Introduction, Matrix Norm and Condition Number, Computation of, Iterative Methods for $Ax = b.$

Linear Algebra and Game Theory **10 Hours**

Linear Inequalities, The Simplex Method, The Dual Problem, Network, GameTheory . .

Text Book:

Linear Algebra and Its Applications by Gilbert Strang, 2016 Edition, Wellesley-Cambridge Press and SIAM, ISBN: 978-09802327-7-6.

Reference Book:

1. Numerical Linear Algebra, William Layton and Myron Sussman, University of Pittsburgh, Pennsylvania, ISBN 978-1-312-32985-0

SDS152 SYSTEM SECURITY

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Identify the security, cryptography techniques and provide security for OS.
CO2: Explain ways of providing security to databases and networks.
CO3: Apply security and privacy techniques to cyber related issues.

Security / Cryptography **10 hours**

The Meaning of Computer Security, Computer Criminals, Methods of Defense. Elementary Cryptography, Terminology and Background, Substitution Ciphers, Transpositions (Permutations), Making a Good Encryption Algorithms, The Data Encryption Standard, The AES Encryption Algorithm, Public Key Encryption, The Uses of Encryption, Summary of Encryption

Program / OS Security **10 hours**

Program Security, Secure Programs, No malicious Program Errors, Viruses and Other Malicious Code, Targeted Malicious Code, Controls Against Program Threats, Summary of Program Threats and Controls
Protection in General-Purpose Operating Systems, Protected Objects and Methods of Protection, Memory and Address Protection, Control of Access to General Objects, File Protection Mechanisms, User Authentication, Summary of Security for Users

Trusted OS / Database Security **10 hours**

Designing Trusted Operating Systems, Security Policies, Models of Security, Trusted Operating System Design, Assurance in Trusted Operating Systems, Summary of Security in Operating Systems
Database and Data Mining Security, Introduction to Database, Security Requirements, Reliability and Integrity, Sensitive Data, Inference, Multilevel Databases, Proposals for Multilevel Security, Data Mining

Network / Administrative Security **10 hours**

Security in Networks, Network Concepts, Threats in Networks, Network Security Controls Section, Firewalls, Intrusion Detection Systems, Secure E-Mail, Summary of Network Security
Administering Security, Security Planning, Risk Analysis, Organizational Security Policies, Physical Security

Cyber Security / Privacy **10 hours**

The Economics of Cyber security, making a Business Case, Quantifying Security, Modeling Cyber security, Current Research and Future Directions, Summary

Privacy in Computing, Privacy Concepts, Privacy Principles and Policies, Authentication and Privacy, Data Mining, Privacy on the Web,. E-Mail Security, Impacts on Emerging Technologies

Text Book:

1. “Security in Computing”, 4th edition, Charles P. Pfleeger - Pfleeger Consulting Group, Shari Lawrence Pfleeger - RAND Corporation

SDS153 EXPLORATORY DATA ANALYSIS AND VISUALIZATION

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

On completion of the course the student should be able to
CO1: Understand the different concepts of EDA techniques
CO2: Explain the techniques for testing EDA assumptions
CO3: Apply data analysis techniques to solve real world problems.
CO4: Conduct a survey for different EDA case studies
CO5: Create visualization using basic and advanced techniques of information visualization and scientific visualization.

Introduction **10 hours**

EDA Introduction, What is EDA? EDA vs Classical & Bayesian, EDA vs Summary, EDA Goals, The Role of Graphics, An EDA/Graphics Example, General Problem Categories.

EDA Assumptions **10 hours**

Underlying Assumptions, Importance, Techniques for Testing Assumptions, Interpretation of 4-Plot, Consequences.

EDA Techniques **10 hours**

Introduction, Analysis Questions, Graphical Techniques: Alphabetical, Graphical Techniques: By Problem Category, Quantitative Techniques, Probability Distributions.

EDA Case Studies **10 hours**

Case Studies Introduction, Case Studies: Normal random numbers, Uniform random numbers, Random walk, Josephson Junction Cry thermometry, Beam Deflections, Filter Transmittance, Standard Resistor, Heat Flow Meter 1, Airplane Glass Failure Time, Ceramic Strength.

Data Visualization **10 hours**

Introduction to R, Rstudio, and Data cleaning and aggregation, Design principles for charts and graphs, ggplot2 and Tableau tools for creating data visualizations, The process creating visualizations and selecting the appropriate visual display, Designing effective digital presentations, Visualization as exploration, Visualizing categorical data, Visualizing time series data, Visualizing multiple variables, Visualizing geospatial data, Dashboard design, Web-based visualizations, Interactive visualizations and motion.

Text Books:

1. Exploratory Data Analysis With R, Roger D.Peng
2. Interactive Data Visualization for the Web, Scott Murra

Reference Books:

1. Engineering Statistics Handbook
2. Advanced Analytics with R and Tableau, Jen Stirrup, Packt Publications

SDS210 ADVANCED DATA MINING TECHNIQUES

No. of Credits: 05 **Total Hours: 50**

Course Outcome:

After completion of this course, students should be able to
CO1:Understand the Data Preparation techniques for Data Mining Process
CO2:Understand the various techniques to Mine the Data Streams
CO3:Exploring Advanced Concepts to build Association among items
CO4:Understand and Explore Data Mining methods as Tools
CO5:Prepare students for research in the area of data mining and related applications

Introduction **10 Hours**

The Data Mining Process: Basic Data Types, The Major Building Blocks: A Bird's Eye View, Scalability Issues and the Streaming Scenario, A Stroll through some Application Scenarios, Data Preparation, Feature Extraction and Portability, **Data Cleaning:** Data Reduction and Transformation, **Similarity and Distances:** Multidimensional Data, Text Similarity Measures, Temporal Similarity Measures, Graph Similarity Measures, Supervised Similarity Functions

Mining Data Stream **10 Hours**

Mining Time-Series Data, Mining Sequence Patterns in Transactional Databases, Mining Sequence Patterns in Biological Data, Graph Mining, Social Network Analysis, Multirelational Data Mining, Multidimensional Analysis and Descriptive Mining of Complex Data Objects, Spatial Data Mining, Multimedia Data Mining, Text Mining, Mining the World Wide Web.

Advanced Concepts in Association Analysis **8 Hours**

Frequent Itemset Generation, Compact Representation of Frequent Itemsets, FP-Growth Algorithms, Handling Categorical and Continuous Attributes, Handling a Concept Hierarchy, Sequntial Patterns, Subgraph Patterns, Infrequent Patterns, Counting Frequent Items in a Stream .

Data Mining Methods as Tools **10 Hours**

Memory-Based Reasoning Methods, Fuzzy Sets in Data Mining, Rough Sets, Support Vector Machines, Genetic Algorithm Support to Data Mining, Performance Evaluation for Predictive Modeling.

Applications and Research Trends in Data Mining 10 Hours

Data Mining Applications (Financial Data Analysis, Retail Industry, Telecommunication Industry, Biological Data Analysis, Other Scientific Applications, Intrusion Detection), Data Mining System Products and Research Prototypes, Statistical Data Mining, Visual and Audio Data Mining, Data Mining and Collaborative Filtering, Data Mining, Privacy, and Data Security, Trends in Data Mining, Present Research Avenues.

Text Books:

1.**Data Mining: Concepts and Techniques,** ***Jiawei Han, MichelineKamber, Jian Pei Professor*****,** Third Edition, Morgan Kauffmann Publishers, 2011.

2.**Advanced Data Mining Techniques,** ***David L. Olson, DursunDelen,*** Springer Publisher, 2008

Reference Books:

1.**Data Mining: The Textbook,** ***Charu C. Aggarwal,*** First Edition, Springer Publisher, 2016.

1. Data Mining: Introductory and Advanced Topics,*Dunham,* **First Edition, Pearson Education India Publisher, 2006.**

SDS220 CLOUD COMPUTING AND VIRTUALIZATION

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

After completing this course, the students would be able to:
CO 1: Explain the various services offered by the cloud.
CO 2: Understand the cloud computing delivery models.
CO 3: Distinguish different types of virtualization.
CO 4: Apply scheduling algorithms for cloud applications.
CO 5: Illustrate various security issues in cloud computing.

Introduction **10 Hours**

Introduction, Cloud Infrastructure Cloud computing, Cloud computing delivery models and services, Ethical issues, Cloud vulnerabilities, Cloud computing at Amazon, Cloud computing the Google perspective, Microsoft Windows Azure and online services, Open-source software platforms for private clouds, Cloud storage diversity and vendor lock-in, Energy use and ecological impact, Service level agreements, User experience and software licensing.

Computing **10 Hours**

Cloud Computing: Application Paradigms. Challenges of cloud computing, Architectural styles of cloud computing, Workflows: Coordination of multiple activities, Coordination based on a state machine model: The Zookeeper, The Map Reduce programming model, A case study: The Grep The Web application , Cloud for science and engineering, High-performance computing on a cloud, Cloud computing for Biology research, Social computing, digital content and cloud computing.

Virtualization **10 Hours**

Cloud Resource Virtualization. Virtualization, Layering and virtualization, Virtual machine monitors, Virtual Machines, Performance and Security Isolation, Full virtualization and paravirtualization, Hardware support for virtualization, Case Study: Xen a VMM based paravirtualization, Optimization of network virtualization, vBlades, Performance comparison of virtual machines, The dark side of virtualization.

Management/Scheduling **10 Hours**

Cloud Resource Management and Scheduling. Policies and mechanisms for resource management, Application of control theory to task scheduling on a cloud, Stability of a two-level resource allocation architecture, Feedback control based on dynamic thresholds, Coordination of specialized autonomic performance managers, A utility-based model for cloud-based Web services, Resourcing bundling: Combinatorial auctions for cloud resources, Scheduling algorithms for computing clouds, Fair queuing, Start-time fair queuing, Borrowed virtual time, Cloud scheduling subject to deadlines, Scheduling Map Reduce applications subject to deadlines, Resource management and

dynamic scaling.

Security **10 Hours**

Cloud Security, Cloud Application Development. Cloud security risks, Security: The top concern for cloud users, Privacy and privacy impact assessment, Trust, Operating system security, Virtual machine Security, Security of virtualization, Security risks posed by shared images, Security risks posed by a management OS, A trusted virtual machine monitor, Amazon web services: EC2 instances, Connecting clients to cloud instances through firewalls, Security rules for application and transport layer protocols in EC2, How to launch an EC2 Linux instance and connect to it, How to use S3 in java, Cloud-based simulation of a distributed trust algorithm, A trust management service, A cloud service for adaptive data streaming, Cloud based optimal FPGA synthesis.

Text Book:

1. Dan C Marinescu: Cloud Computing Theory and Practice. Elsevier(MK) 2013.

Reference Books:

1. RajkumarBuyya , James Broberg, AndrzejGoscinski: Cloud Computing Principles and Paradigms, Willey 2014.
2. John W Rittinghouse, James F Ransome:Cloud Computing Implementation, Management and Security, CRC Press 2013.

SDS230 DEEP LEARNING

Total Teaching hours: 50 **Credits: 5**

Course Outcomes:

After completion of the course student should be able to:
CO1: Explain the basic concepts in the design of neural networks
CO2: Apply the concepts of neural networks in designing deep learning applications
CO3: Illustrate parameter tuning for deep learning architecture

Introduction to shallow network **11Hrs**

A review of machine learning: how can machines learn, fundamentals for making a machinelearning; linear algebra, statistics, and simple learning models: regression, clustering, classification. Foundations of neural networks: an introduction to the neural network, training neural network, activation function, loss function, hyperparameters.

Introduction to deep learning **12Hrs**

Fundamentals of deep learning: def, architecture, building blocks of deep learning. Major architecture of deep learning: unsupervised pertained network, convolution neural network, recurrent neural network, recursive neural network.

Building a simple neural network **12Hrs**

Building a neural network: matching neural network for the right problem, DL4J suit tools, modeling CSV data for MLP, modeling handwritten images using CNN, modeling sequence data using RNN, using autoencoders for anomaly detection, using variational autoencoders for reconstructing MNIST data.

Tuning the deep neural network **15Hrs**

Tuning deep neural network: matching input data and network architecture, relating model goals and network parameters, working with layer count, parameter count and memory, weight initialization strategies, apply loss function, applying optimization, controlling epochs and batch size, dealing with overfitting. Tuning specific deep network architecture: CNN, RNN, Restricted Boltzmann machine, DBN. Vectorization: an introduction to vectorization in machine learning, using dataVec for ETL, vectorizing image data, working with sequential data and vectorization, working with text data.

Text Book:

1. Josh Patterson and Adam Gibson, Deep Learning A Practitioner's Approach, Oreilly,2017.

Reference Books:

1. Nikhil Buduma, "Fundamentals of Deep Learning", Oreilly,2017
2. Ian Goodfellow, YoshuaBengio, Aaron Courville., "Deep Learning", 2016,The MIT Press, Cambridge, Massachusetts, London.

SDS241 OPTIMIZATION THEORY

Total Teaching Hours: 50 **No. of Credits : 05**

Course outcomes:

On successful completion of the course, students should be able to:
CO1: Apply classical optimization techniques and modern methods.
CO2: Develop linear and Non-linear programming.
CO3: Identify optimized solutions to constrained systems
CO4: Contrast single variable and multivariable optimization problems.
CO5: Analyse different system and evaluate solution to optimize resources.

Introduction **10 Hours**

Introduction to Optimization: Introduction, Historical Development, Engineering Applications of Optimization' Statement of an Optimization Problem, Classification of Optimization Problems.

Classical Optimization Techniques: Single-Variable Optimization, Multivariable Optimization with No Constraints, Multivariable Optimization with Equality Constraints, Multivariable Optimization with Inequality Constraints, Convex Programming Problem.

Linear Programming **10 Hours**

Applications of Linear Programming, Standard Form of a Linear Programming Problem, Geometry of Linear Programming Problems, Definitions and Theorems, Solution of a System of Linear Simultaneous Equations, Pivotal Reduction of a General System of Equations, Motivation of the Simplex Method, Simplex Algorithm, Improving a Nonoptimal Basic Feasible Solution, Two Phases of the Simplex Method, Revised Simplex Method, Duality in Linear Programming.

Nonlinear Programming **10 Hours**

Introduction, Unimodal Function, Elimination Methods: Unrestricted Search, Exhaustive Search, Dichotomous Search, Interval Halving Method Fibonacci Method, Golden Section Method, Comparison of Elimination Methods. Interpolation Methods: Quadratic Interpolation Method, Cubic Interpolation Method, Direct Root Methods. Direct Search Methods: Random Search Methods, Grid Search Method, Univariate Method, Pattern Directions, Powell’s Method.

Geometric Programming **10 Hours**

Posynomial, Unconstrained Minimization Problem, Solution of an Unconstrained Geometric Programming Program Using Differential Calculus, Solution of an Unconstrained Geometric Programming Problem Using Arithmetic–Geometric Inequality, Primal–Dual Relationship and Sufficiency Conditions in the Unconstrained Case, Constrained Minimization, Solution of a Constrained Geometric Programming Problem, Primal and Dual Programs in the Case of Less-Than Inequalities, Geometric

Programming with Mixed Inequality Constraints, Complementary Geometric Programming.

Stochastic Programming: Basic Concepts of Probability Theory, Stochastic Linear Programming, Stochastic Nonlinear Programming.

Modern Methods **10 Hours**

Genetic Algorithms, Simulated Annealing, Particle Swarm Optimization, Ant Colony Optimization, Optimization of Fuzzy Systems, Neural-Network-Based Optimization.

Text Book:

1. Singiresu S Rao, “Engineering Optimization Theory and Practice", John Wiley and sons, 4th Edition 2009.

Reference Book:

1. Edwin K. P. Chong and Stanislaw. Zak “An Introduction to Optimization”, John Wiley and sons, 2nd Edition 2001.

SSE242 COMPUTATIONAL LINGUISTICS

Total Teaching Hours: 50 **No. of Credits : 05**

Course outcomes

On successful completion of the course, students should be able to:

CO1: Explain finite state transducers for language processing, rules and analysis.
CO2: Demonstrate parsers in computational linguistics
CO3: Implement natural language processing applications.
CO4: Discuss clustering of text data in natural language processing applications.
CO5: Illustrate machine translation system using machine learning algorithms.

Introduction **10 Hours**

What is computational linguistics? Ambiguity and uncertainty in language, regular languages, and their limitations, finite-state automata, morphology. Natural Language Toolkit.

Context Free Grammars **10 Hours**

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

Computational Discourse **10 Hours**

Discourse segmentation, Text coherence, reference resolution, reference phenomena, Features for pronominal anaphora resolution, Algorithms for anaphora resolution, coreference resolution, evaluation of coreference resolution.

Word Sense Disambiguation and Clustering **10 Hours**

Holonomy, polysemy, different meanings, the power of context. Language neighborhood as a vector. Agglomerative clustering. Clustering by expectation maximization. Using clustering to discover different word senses. Semi-supervised document classification.

Machine Translation **10 Hours**

Probabilistic models for machine translation system, alignment, translation, language generation. Machine translation evaluation.

Text Books:

1. Daniel Jurafsky and James H. SPEECH and LANGUAGE PROCESSING: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Second Edition.
2. Chris Manning and HinrichSchütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.

SDS243 BIOINFORMATICS

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcome :

On successful completion of the course students will have the ability to:

CO1 : Explain Bioinformatics, biological databases and data models.
CO2: Obtain sequence data from biological databases and compare different file formats.
CO3 : Analyze and evaluate methods of sequence alignment, utilize similarity finding tools to compare sequence data.
CO4 : Design applications using bioinformatics algorithms for gene finding, predictions and genome studies.
CO5 : Engage in independent study to analyze and interpret protein structure using biostatistics & biological tools and compare different methods and strategies of structure prediction.

Introduction and NCBI Data Model **8 Hours**

Introduction to Bioinformatics, Goal, Scope, Applications, Limitations, New Themes, Introduction to Database, Types of Databases, Biological Databases, Pitfalls of Biological Databases. Introduction to NCBI data model, PUBs: Publications or Perish, SEQ-Ids: What's in a Name? BIOSEQs: Sequences, BIOSEQ-SETs: Collections of Sequences, SEQ-ANNOT: Annotating the Sequence,SEQ-DESCR: Describing the Sequence, Using the Model.

Bioinformatics Database **10 Hours**

Importance of Databases, Characteristics and Categories of Bioinformatics Database, Navigating Databases, Biological Databases, Primary Sequence Databases, Composite Sequence Databases, Secondary Databases, Nucleic Acid Sequence Databases, Structure Databases: File Formats, Protein Structure, PDB, MMDB, CATH, Other Database Enzyme, MEROPS, BRENDA, Pathway databases, Bibliographic Databases, Specialized Genomic Resources, Analysis Packages.

Sequence Align Methods **10 Hours**

Sequence Analysis of Biological Data, Significance of Sequence Alignment, Pairwise Sequence Alignment Methods, Use of Scoring Matrices and Gap Penalties in Sequence Alignments, Multiple Sequence Alignment Methods - Tools and Application of multiple sequence alignment, Gene Prediction Strategies, Protein Prediction Strategies, Phylogenetic Trees and Multiple Alignments.

Bioinformatics Algorithms **12 Hours**

Biological Algorithms versus Computer Algorithms, Exhaustive Search, Mapping Algorithms, Motif Finding Problem, Search Trees, Finding a Median String, Greedy Approach to Motif Finding, DNA Sequence comparison - Manhattan Tourist Problem - Edit Distance and Alignments - Longest Common Subsequences - Global Sequence Alignment - Scoring Alignment - Local Sequence Alignment – Alignment with Gap Penalties - Multiple Alignment, DNA Sequencing, Shortest Superstring Problem, DNA arrays as an alternative sequencing techniques.

Biostatistics & Tools **10 Hours**

Handling Univariate and Bivariate Data, Measures of Central Tendency, Measures of Dispersion, Skewness & Kurtosis, Correlation and Regression.
Local Alignment Search Tool (BLAST), Purpose of BLAST, BLAST Analysis, Purpose of BLAST II, Scoring Metrics, PAM, BLOSUM, Working of BLAST, Introduction to HMMER.

Text Books:

1. JinXiong (2006) Essential Bioinformatics, Cambridge University Press,
2. Baxevanis A. D. and B. F. Francis Ouellette, (2001) Bioinformatics a practical guide to the analysis of genes and proteins. Second edition, John Wiley and Sons.
3. An Introduction to Bioinformatics Algorithms, Neil C Jones and Pavel A Pevzner, MIT Press, 2004.

Reference Books:

1. Bioinformatics - Concepts, Skills, and Applications, S.C. Rastogi, NamitaMendiratta, ParagRastogi, Second Edition, CBS Publishers, 2003.
2. Bioinformatics: Databases, Tools, And Algorithms.,OrpitaBosu, Simminder Kaur Thukral, Oxford University Press Publisher, 2007.
3. Fundamentals of Mathematical Statistics., S.C. Gupta and V.K. Kapoor, Eleventh Edition, Sultan Chand & Sons Publishers, 2007.

Online Courses:

1. NPTEL: https://nptel.ac.in/courses/102106065/

SDS251 SCALABLE SYSTEMS FOR DATA SCIENCE

Total Teaching Hours: 50 **No. of Credits : 05**

Course outcomes:

At the end of the course, students will have the ability to:

CO1: Explain big data platforms and distributed systems.
CO2: Obtain large scale of input data and output results using different algorithmic techniques and software paradigms.
CO3: Analyze algorithms and modify to adapt to a distributed setting.
CO4: Design and develop scalable algorithms and systems for the common data science tasks.
CO5: Engage in independent study and make an effective oral presentation on applications of Scalable data science.

Big Data **10 hours**

Big Data & Platform Design Goals, Big Data & other computing platforms. Programming for Large Datasets:. Distributed systems, scalability and metrics, Degrees of Parallelism, MapReduce- Uses, Model working, simple and advanced applications programming. Runtime Systems: Hadoop- Open cloud server, Class cluster, Hadoop distributed file system. Hadoop YARN, Hadoop MapReduce, Fault Tolerance.

Prediction **10 hours**

Prediction over graphs – Scalable learning and inference over graphs, Semi supervised learning (SSL)- self training and co training, Graph Based SSL. Streaming Naive Bayes – Introduction to Naïve Bayes, Complexity of Naïve Bayes, Implementation of Naïve Bayes Classifier, Large vocabulary counting, Sorting – Merge sort, Unix sort, Large vocabulary Naïve Bayes, Distributed Counting, optimizations.

Learning **10 hours**

Scalable Logistic Regression and SGD. Learning as optimization, Stochastic gradient descent, SGD versus streaming, Logistic regression versus Rocchio and Naïve Bayes, Efficient Logistic Regression with Stochastic Gradient Descent, Regularized logistic regression, Sparse updates for Regularized logistic regression, Bounded memory logistic regression, SGD implementation.

Matrix Factorization **10 hours**

Large-scale Matrix Factorization (MF) - Recovering Latent factor in a matrix, Matrix factorization for collaborative filtering, MF for image and text modeling, Large scale MF for distributed SGD, Distributed SGD for Mapreduce.MR Advanced Topics: Inverted Index, PageRank, Dist

ributed graph processing: Apache Giraph, GoFFish.

Distributed stream processing **10 hours**

Distributed and fault-tolerant realtime computation, Distributed Stream Processing systems- Apache Storm. Parameter Server- Architecture, Key- value vectors, Range Push and Pull, User defined functions on the server, Asynchronous tasks and dependency, Flexible consistency, user defined filters, messages, consistent hashing, server management, worker management. Evaluation of parameter server - Sparse Logistic Regression and Latent Dirichlet Allocation.

Text Books:

1. Select chapters from Mining of Massive Datasets, JureLeskovec, AnandRajaraman and Jeff Ullman, 2nd Edition (v2.1), 2014.
2. Select chapters from Data-Intensive Text Processing with MapReduce, Jimmy Lin and Chris Dyer, 1st Edition, Morgan & Claypool Publishers, 2010

References:

1. Research papers and articles - [MR for ML on Multicore, NIPS 06], [Hogwild!], [Bottou, 2010], [Gemulla et al., KDD 2011]

Online courses:

1. NPTEL: https://nptel.ac.in/courses/106105186/
2. SWAYAM: https://swayam.gov.in/nd1_noc19_cs61/preview

SDS252 WEB DATABASES AND INFORMATION SYSTEMS

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

After completion of course students should be able to:
CO1: Illustrate the concepts of web-based information systems.
CO2: Explain n-tiered architectures to implement secure, scalable systems
CO3: Demonstrate database driven applications.
CO4: Design and develop web-based applications.
CO5: Describe XML and alternative technologies.

Web-based Information system **10 hours**

Web -Based Information Systems, Applications: electronic commerce, Variants of Web database access, Basic Web Standards, architecture. HTTP, Forms, Server-Side Programming and CGI Alternatives to CGI: Java Servlets and server APIs Browser detection, state, cookies and redirects

Using n-tiered architectures to implement secure and scalable systems 10 hours

Web App Architectures: Multi-Tier (2-Tier, 3-Tier) Model-Viewer-Controller (MVC)

Database-driven websites and applications **10 Hours**

Utilize JavaScript to improve database driven websites. − Critical components of the modern Web infrastructure: DNS, CDN, etc Critical components of the modern Web infrastructure: DNS, CDN, etc

DBMS and WWW: **10 hours**

Introduction Off-Line access to databases Static and Dynamic Web Pages SQL embedded in HTML CGI solution to database gateway Internet database connector JDBC: databases the Java way. Solutions from database vendors Association rule mining

XML and its alternatives **10 Hours**

HTTP, XML, SQL, JavaScript, AJAX XML and its Alternatives: XML: Basics of XML, namespace, schema languages, XSLT and XPath, alternatives to XML, SQL, CSS, RSS, and others.

Text Books:

1. PROFESSIONAL WEB 2.0 PROGRAMMING: USING XHTML, CSS, JAVASCRIPT AND AJAXBy Eric Van Der, Danny Ayers, Erik Bruchez

2. Weaving a Website - Programming in HTML, Javascript, Perl, and Java

Reference Book:

1. Mastering HTML, CSS &Javascript Web Publishing Paperback – 15 Jul 2016 by Laura Lemay, Rafe Colburn, JenniferKyrnin.

SDS253 SOCIAL AND INFORMATION NETWORK ANALYSIS

Total Teaching Hours: 50 **No. of Credits : 05**

Course Outcomes:

After completion of course students should be able to:
CO1: Analyzethe social and information network.
CO2: Apply methods to evaluate the behavior of cascading networks.
CO3: Design a graph model representing a social network.
CO4: Apply different search algorithms over the graph model representing social network.
CO5: Interpret the diffusion over social and information network.

Introduction **10 hours**

Overview: Aspects of Networks, Central Themes and Topics. Graphs**:** Basic Definitions, Paths and Connectivity, Distance and Breadth-First Search, Network Datasets: An Overview

The Small-World Phenomenon **10 hours**

Six Degrees of Separation, Structure and Randomness, Decentralized Search, Empirical Analysis and Generalized Models, Core-Periphery Structures and Difficulties in Decentralized Search, Advanced Material: Analysis of Decentralized Search.
Positive and Negative Relationships: Structural Balance , Characterizing the Structure of Balanced Networks, Applications of Structural Balance , A Weaker Form of Structural Balance , Advanced Material: Generalizing the Definition of Structural Balance.

Cascading Behaviour in Networks **10 hours**

Diffusion in Networks, Modeling Diffusion through a Network, Cascades and Clusters, Diffusion, Thresholds, and the Role of Weak Ties, Extensions of the Basic Cascade Model, Knowledge, Thresholds, and Collective Action, Advanced Material: The Cascade Capacity.
Epidemics : Diseases and the Networks that Transmit Them , Branching Processes, The SIR Epidemic Model , The SIS Epidemic Model , Synchronization , Transient Contacts and the Dangers of Concurrency , Genealogy, Genetic Inheritance, and Mitochondrial Eve, Advanced Material: Analysis of Branching and Coalescent Processes

Power Laws and Rich-Get-Richer Phenomena **10 hours**

Popularity as a Network Phenomenon, Power Laws, Rich-Get-Richer Models, The Unpredictability of Rich-Get-Richer Effects, The Long Tail, The Effect of Search Tools and Recommendation Systems Advanced Material: Analysis of Rich-Get-Richer Processes.
The structure of the Web : The World Wide Web, Information Networks, Hypertext,

and Associative Memory The Web as a Directed Graph, The Bow-Tie Structure of the Web, The Emergence of Web 2.0 U

Link Analysis and Web Search **10 hours**

Searching the Web: The Problem of Ranking , Link Analysis using Hubs and Authorities, PageRank, Applying Link Analysis in Modern Web Search, Applications beyond the Web, Advanced Material: Spectral Analysis, Random Walks, and Web Search

Strong and Weak Ties: Triadic Closure , The Strength of Weak Ties, Tie Strength and Network Structure in Large-Scale Data , Tie Strength, Social Media, and Passive Engagement Closure, Structural Holes, and Social Capital, Advanced Material: Betweenness Measures and Graph Partitioning

Text Book:

1. “Networks, Crowds, and Markets Reasoning about a Highly Connected World”, David Easley, Cornell University, New York, Jon Kleinberg, Cornell University, New York, 2010

Reference Books:

1. “Networks: An Introduction By M. E. J. Newman, A College-Level Textbook About The Science Of Networks.”, M. E. J. Newman Hardback, Oxford University Press, 2010