MANIPAL UNIVERSITY
JAIPUR

SCHOOL OF COMPUTING & INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION TECHNOLOGY

MASTER OF TECHNOLOGY (DATA SCIENCES)

COURSE STRUCTURE & SYLLABI

(For Batch: 2021-23)

School of Computing and Information Technology
Department of Information Technology

Course Structure

For M. Tech (Data Sciences) Program from Academic Session: 2019-20

Duration: 2 Years, Total Number of Credits: 75

Year	FIRST SEMESTER						SECOND SEMESTER					
	Sub. Code	Subject Name	L	T	P	C	Sub. Code	Subject Name	L	T	P	C
I	IT6101	Mathematical Foundations of Data Science	3	1	0	4	IT6201	Big Data Analytics	3	1	0	4
	IT6102	Advanced Data Structures and Algorithms	3	1	0	4	IT6202	Deep Learning	3	1	0	4
	IT6103	Machine Learning Techniques	3	1	0	4	IT6203	Data Visualization	3	1	0	4
	IT61XX	Program-Elective-I	3	0	0	3	IT62XX	Program-Elective-III	3	0	0	3
	IT61XX	Program-Elective-II	3	0	0	3	IT62XX	Program-Elective-IV	3	0	0	3
	DR6001	Research Methodology	3	0	0	3	IT6204	Research Practice	0	0	6	3
	IT6130	Advanced Data Structures and Algorithms Lab	0	0	4	2	IT6230	Big Data Analytics Lab	0	0	4	2
	IT6131	Machine Learning Techniques Lab	0	0	4	2	IT6231	Deep Learning Lab	0	0	2	1
							IT6232	Data Science Project Lab	0	0	2	1
	Total Contact Hours (L + T + P)		18	3	8	25	Total Contact Hours (L + T + P)		15	3	14	25
	Program Elective-I Choices • IT6140 Distributed Systems • IT6141 Data Preparation and Analysis **Program Elective-II Choices** • IT6142 Storage Area Networks • IT6143 Knowledge Discovery						**Program Elective-III Choices** • IT6242 Image Processing and Pattern Recognition • IT6243 Recommender Systems **Program Elective-IV Choices** • IT6244 Web Analytics and Development • IT6245 Data Access Control and Security					
	Total Contact Hours (L + T + P) = 29 Total Credits: 25						Total Contact Hours (L + T + P) = 32 Total Credits = 25					
II	**THIRD SEMESTER**						**FOURTH SEMESTER**					

	IT7170 Dissertation-I / IT7171 IndustryTraining-I	0	0	24	12	IT7270 Dissertation-II / IT7271 IndustryTraining-II	0	0	26	13
	Total Contact Hours = 24 Total Credits: 12					Total Contact Hours = 26 Total Credits: 13				

IT6101: MATHEMATICAL FOUNDATIONS OF DATA SCIENCE [3 1 0 4]

Basics of Data Science: Introduction, typology of problems, importance of linear algebra, statistics and optimization from a data science perspective, structured thinking for solving data science problems; **Linear Algebra:** Matrices and their properties (determinants, traces, rank, nullity, etc.), eigenvalues and eigenvectors, Matrix factorizations, inner products, distance measures, projections, notion of hyper planes, half-planes; **Probability, Statistics and Random Processes:** Probability theory and axioms, random variables, probability distributions and density functions (Uni-variate and multivariate), expectations and moments, covariance and correlation, statistics and sampling distributions, hypothesis testing of means, proportions, variances and correlations, confidence (statistical) intervals, correlation functions, white-noise process; **Optimization:** Unconstrained optimization, necessary and sufficiency conditions for optima, gradient descent methods, constrained optimization, KKT conditions, introduction to non-gradient techniques, introduction to least squares optimization, optimization view of machine learning; **Introduction to Data Science Methods:** Linear regression as an exemplar function approximation problem, linear classification problems.

References:

1. G. Strang *Introduction to Linear Algebra*, (5e), Wellesley-Cambridge Press,2016.
2. Bendat, J. S. and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, (4e), John Wiley & Sons, 2010.
3. Montgomery, D. C. and G. C. Runger, *Applied Statistics and Probability for Engineers*, (5e), John Wiley & Sons, 2011. 4. Cathy O'Neil and Rachel Schutt, *Doing Data Science*, (4e), O'Reilly Media, Fourth Edition, 2016.

IT6102: ADVANCED DATA STRUCTURES AND ALGORITHMS [3 1 0 4]

Advanced Search Trees: Review of Binary Search Trees, AVL Tree, R-B Trees and Splay Trees, Advanced Search Data Structures Like- Treaps, Skip Lists, Finger Search Trees, Biased Search Trees; **Data Structures For External Storage:** Review of 2-3-4 Trees and 2-3 Trees, B-Tree, B+ Trees, Priority Queues and Concatenable Queues Using 2-3 Trees; **Advanced Heaps:** Review of Heaps, Binomial Trees, Implementing Binomial Heaps and its Operations, Structure of Fibonacci Heaps, Mergeable Heap
Operations, Decreasing Key and Deleting a Node, Bounding the Maximum Degree, Amortized Analysis of
Fibonacci Heaps; **Dictionaries and Hashing:** Review of Dictionaries and Implementation, Review of Hashing- The Bucket Approach, Index File Approach, Universal Hashing, Perfect Hashing, LocalitySensitive Hashing, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing, Synopses, Fingerprints, Fault Tolerant Data Structures; **Graph Theory And Sorting Network**: Review Of Graph Representation and Basic Algorithms, Algorithms for Connectedness, Finding all Spanning Trees in a Weighted Graph, Bipartite
Graphs: Maximum Matching, the Hungarian Algorithm, Maximum Flow in a Transport Network - the Ford– Fulkerson Algorithm; **Sorting Network:** Comparison Network, Zero-One Principle, Bitonic Sorting and Merging Network Sorter.

References:

1. Cormen, T. H., Leiserson, C. E., Rivest, R.L., and Stein, C. *Introduction to Algorithms*, (3e), MIT Press, 2009.
2. Aho, V., Hopcroft, J. E. and Ullman, J. D. *The Design and Analysis of Computer Algorithms*, (1e), Fourth Impression, Pearson Education, 2009.
3. Horowitz, E., Sahni, S. and Rajasekaran, S, *Computer Algorithms*, (2e), University Press, 2007.
4. Weiss, M. A, *Data Structures and Algorithm Analysis in C++*, (2e), Pearson Education India, 2004.
5. Goodrich, M. T., Tamassia, R., *Algorithm Design*, (1e), John Wiley, 2002.

IT6103: MACHINE LEARNING TECHNIQUES [3 1 0 4]

Introduction: learning problems, perspectives and issues, concept learning, version spaces and candidate eliminations, inductive bias, decision tree learning, representation, algorithm, heuristic space search. Linear and logistic regression; **Neural Networks and Genetic Algorithms:** Neural network representation, problems, perceptron's, multilayer networks and back propagation algorithms, advanced topics, hyper parameter optimization, genetic algorithms, hypothesis space search, genetic programming, models of evaluation and learning; **Bayesian and Computational Learning:** Bayes theorem, concept learning, maximum likelihood – minimum description length principle, Bayes optimal classifier, Gibbs algorithm, naïve Bayes classifier, Bayesian belief network, EM algorithm, probability learning, sample complexity, finite and infinite hypothesis spaces, mistake bound model; **Instant Based Learning**: K- nearest neighbor learning, locally weighted regression, radial basis functions, case based learning; **Advanced Learning:** Learning sets of rules, sequential covering algorithm, learning rule set, first order rules, sets of first order rules, induction on inverted deduction, inverting resolution, analytical learning, perfect domain theories, explanation base learning, FOCL algorithm, reinforcement learning task, Q-learning, temporal difference learning.

References:

1. Tom M. Mitchell, *Machine Learning*, (2e), McGraw-Hill Education India Private Limited, 2017.
2. Ethem Alpaydin, *Introduction to Machine Learning*, (3e), The MIT Press, 2015.
3. Stephen Marsland, *Machine Learning: An Algorithmic Perspective*, (2e), CRC Press, 2014.

Annexure 2

DR6001: Research Methodology [3 0 0 3]

Introduction to Research: Meaning and Concepts, Types of Research, Process of Research, Classification of Variables, Review of Literature, Formulating Research Problem & determining research objectives, Ethics in Research, Limitations in Research. Research Repositories. Research Design: Meaning & concept of research design, Types of Research Design, Need for research design, Features of a Good research Design, Developing a Research Plan. Data Collection: Types of Data – Primary and Secondary Data – Benefits and Drawbacks, Various methods of Data Collection, Measurement Scales – Classification of Scales, Measurement error, Criteria for Good Measurement. Interpretation and Report Writing: Meaning of Interpretation, Technique of Interpretation, Precaution in Interpretation, Types of Reports, Significance of Report Writing, Different Steps in Writing Report, Layout of the Research Report, Oral Presentation, Thesis writing, research paper writing, preparing synopsis & summary of research thesis work. Publishing research papers, reference writing: foot note, Various electronic tools for citation and referencing, in-text citation, bibliography, citation styles, Bibliometrics. Sampling Techniques: basic terms, Importance of sampling in research, essentials of a good sample, sampling error, standard error of the mean (Standard Deviation), Estimation of parameters, accuracy & precision of estimation, sampling procedure, types/methods of sampling, Central limit theorem, sample size determination, confidence interval and Confidence level. Measurement & Scaling Techniques: - types of data: Primary & Secondary, Types of Scales: Ratio, Interval, Ordinal Nominal. Mapping rules, characteristics of a good measurement, sources of error in measurement.

References:

1. C R Kothari and G. Garg, *Research Methodology: methods and techniques*, New Age International (P) Ltd, 2019.
2. K. S. Bordens and B. B. Abbitt, *Research Design & Methods, A process approach*, (10e), McGraw Hill Education, 2018.
3. D. Chawla and N. Sondhi, *Research Methodology – Concepts and Cases*, Vikas Publishing House, 2011.
4. D. R. Cooper, P. S. Schindler, *Business Research Methods*, (12e), McGraw Hill, 2014.
5. A. Bryman and E. Bell, *Business Research Methods*, (5e), Oxford University Press, 2018.
6. W. G. Zikmund, *Business Research Methods*, (8e), Cenagage Learning, 2010.
7. W. Guddard and S. Melville, *Research Methodology an introduction for Science & Engineering Students*, (2e), Kenwyn, South Africa: Juta & Co., 2011.
8. R. Ganesen, *Research Methodology for Engineers*, MJP Publishers, Chennai, 2011.

IT6130: ADVANCED DATA STRUCTURES & ALGORITHMS LAB [0 0 4 2]

Application of Search Trees: Search, insert and delete operations and their comparison for binary search tree, R-B Trees and 2-3 trees. Analysis of operations on structures studied like skip-lists; **Application of Heaps:** Converting a set of data records in to Binomial heaps and Fibonacci heaps. Comparing the search, add and delete operations based on time and space requirements; **Application of Graph:** Maximum flow problem should be solved using Ford Fulkerson method. A mapping solved using Hungarian Algorithm and Ford Fulkerson algorithm; **Lab Project:** A generalized case study where student will choose appropriate data-structure(s) and apply based on their utility for specific problem.

References:

1. Cormen, T. H., Leiserson, C. E., Rivest, R.L., and Stein, C. *Introduction to Algorithms*, (3e), MIT Press, 2009.
2. Aho, V., Hopcroft, J. E. and Ullman, J. D, *The Design and Analysis of Computer Algorithms*, (1e), Fourth Impression, Pearson Education India, 2009.
3. Horowitz, E., Sahni, S. and Rajasekaran, S, *Computer Algorithms*, (2e), University Press, 2007.
4. Weiss, M. A, *Data Structures and Algorithm Analysis in C++*, (2e), Pearson, 2004.
5. Goodrich, M. T., Tamassia, R., *Algorithm Design*, (1e), John Wiley, 2002.

IT6131: MACHINE LEARNING TECHNIQUES LAB [0 0 4 2]

Classification, Clustering Algorithms, Naïve bayes classifier, Neural Network related experiments and programs, Performance evaluation, case study

1. Tom M. Mitchell, Machine Learning, (2e), McGraw-Hill Education India Private Limited, 2017.
2. Ethem Alpaydin, Introduction to Machine Learning, (3e), The MIT Press, 2015.

IT6201: BIG DATA ANALYTICS [3 1 0 4]

Introduction to Big Data: Introduction, Distributed File System, Big Data and its Importance, Drivers, Big Data Analytics, Big Data Applications, Algorithms, Matrix-Vector Multiplication by MapReduce; **Big Data Analytics:** Analyzing Big Data, Sources of Big Data, Characteristics of Big Data, Drivers of BDA, Types of Data, Structured vs. Unstructured Data, Data Marts, Case Study Based Tutorial, Differences Between Traditional DWDM and BDA, Limitations of Traditional RDBMS to Store and Analyses Big Data, Data science, Definition and Concepts; **Data Scientists:** Key Competencies and Characteristics of Data Scientists, More Discussions on Data Science, Data Wrangling, Data Mugging, Data Jujitsu, Tutorial Based on Data Science Applications, Big Data Analytics Ecosystem; **State of the Practice in Analytics:** Data Analytics Lifecycle and Discussions, Roles for a Successful Analytics Project, Case Study to Apply the Data Analytics Lifecycle, Analytical Databases and DW Appliances, Hadoop Distributions – Comparing Various BDA Tools; **Analyzing and Exploring the Data:** Challenges when Managing and Analyzing Big Data, The Role of Data Virtualization in a Big Data Environment; **Big Data Platforms and Storage Options:** The New Multi-Platform Analytical Ecosystem, Beyond the Data Warehouse - Analytical Databases, Hadoop and NoSQL DBMS; **Introduction to HADOOP:** Big Data, Apache Hadoop &Hadoop Ecosystem, Map Reduce, Data Serialization; **HADOOP Architecture:** Architecture, Storage, Task trackers, HadoopConfiguration HADOOP; **Ecosystem and yarn:** Hadoop Ecosystem Components, Hadoop 2.0 New Features NameNodeHigh Availability, HDFS Federation, MRv2, YARN, and Running MRv1 in YARN; **Spark, Hive and Pig**: Introduction and Applications.

References

1. EMC Education Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, (1e), John Wiley & Sons, 2015.
2. Minelli, Michael, Michele Chambers, and Ambiga Dhiraj. *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*, (1e), John Wiley & Sons, 2012.
3. Bahga, Arshdeep and Vijay Madisetti, *Big data science & analytics: A hands-on approach*, (1e), VPT, 2016.

IT6202: DEEP LEARNING [3 1 0 4]

Introduction to Deep Learning: history of deep learning, deep learning success stories, mcculloch pitts neuron, thresholding logic, perceptron's, perceptron learning algorithm; **Multi-Layer Network and Optimization Technique:** multilayer perceptron's (mlps), representation power of mlps, sigmoid neurons, gradient descent, feed forward neural networks, representation power of feed forward neural networks feed forward neural networks, back propagation gradient descent (gd), momentum based gd, nesterov accelerated gd, stochastic gd, adagrad, rmsprop, adam, eigenvalues and eigenvectors, eigenvalue decomposition, basis; **Dimension Reduction and Regularization:** principal component analysis and its interpretations, singular value decomposition auto encoders and relation to pca, regularization in auto encoders, denoising auto encoders, sparse auto encoders, contractive auto encoders regularization: bias variance tradeoff, l2 regularization, early stopping, dataset augmentation, parameter sharing and tying, injecting noise at input, ensemble methods, dropout greedy layer wise pre-training, better activation functions, better weight initialization methods, batch normalization learning vectorial representations of words; **Convolutional Neural Networks:** lenet, alexnet, zf-net, vggnet, googlenet, resnet, visualizing convolutional neural networks, guided back propagation, deep dream, deep art, fooling convolutional neural networks; **Recurrent Neural Networks:** back propagation through time (bptt), vanishing and exploding gradients, truncated bptt, gru, lstms encoder decoder models, attention mechanism, attention over images.

References:

1. J.Patterson, A.Gibson, *Deep Learning*, (1e), O'Reilly Publication, 2018.
2. Goodfellow I., Bengio Y, *Deep Learning (Adaptive Computation and Machine Learning series)*, (1e), MIT Press, 2017.
3. Shai Shalev-Shwartz , Shai Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, (3e), Cambridge University Press, 2015.

IT6203: DATA VISUALISATION [3 1 0 4]

Introduction: Visual Representation of Data, Gestalt Principles, Information Overloads; **Creating Visual Representations**: Visualization Reference Model, Visual Mapping, Visual Analytics, Design of Visualization Applications; **Classification of Visualization Systems**: Interaction and Visualization Techniques Misleading, Visualization of One, Two and Multi-Dimensional Data, Text and Text Documents; **Visualization of Groups:** Trees, Graphs, Clusters, Networks, Software, Metaphorical Visualization; **Visualization of Volumetric Data:** Vector Fields, Processes and Simulations, Visualization of Maps, Geographic Information, GIS systems, Collaborative Visualizations, Evaluating Visualizations; **Recent Trends in Various Perception Techniques:** Various Visualization Techniques, Data Structures used in Data Visualization.

References:

1. Matthew Ward and Georges Grinstein, *Interactive Data Visualization: Foundations, Techniques, and Applications*, (2e), A K Peters/CRC Press, 2015.
2. Jurgen Kai-Uwe Brock, *Data Design: The Visual Display of Qualitative and Quantitative Information*, (1e), Consulting Press, 2017.
3. Edward R. Tufte, *The Visual Display of Quantitative Information*, (2e), Graphics Press USA, 2001.
4. Cole Nussbaumer Knaflic, *Storytelling With Data: A Data Visualization Guide for Business Professionals*, (1e), John Wiley and Sons, 2015.

IT6204: RESEARCH PRACTICE [0 0 6 3]

The aim of the course is to give training and opportunities to the post-graduate students in gaining skills and competence in research methodologies and writing research reports / proposals. The course has to be completed by the student under the guidance of a supervisor. The supervisor would assign to student an appropriate research oriented problem to work which is related to his own domain of expertise. The course is a three unit course and a student is expected to do six hours of research oriented development activity.

IT6230: BIG DATA ANALYTICS LAB [0 0 4 2]

Introduction to HADOOP: Big Data, Apache Hadoop & Hadoop Ecosystem, Map Reduce, Data Serialization; **HADOOP Architecture:** Architecture, Storage, Task trackers, Hadoop Configuration HADOOP**; Ecosystem and Yarn:** Hadoop Ecosystem Components, Hadoop 2.0 New Features Name Node High Availability, HDFS Federation, MRv2, YARN, and Running MRv1 in YARN; **Spark, Hive and Pig**: Introduction and Applications.

References:

1. EMC Education Services, *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*, (1e), John Wiley & Sons, 2015.
2. Minelli, Michael, Michele Chambers, and Ambiga Dhiraj. *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*, (1e), John Wiley & Sons, 2012.
3. Bahga, Arshdeep and Vijay Madisetti, *Big data science & analytics: A hands-on approach*, (1e), VPT, 2016.

IT6140: DISTRIBUTED SYSTEMS [3 0 0 3]

Introduction: Definition, Goals, Types of Distributed Systems and Architectural Styles. Communication: Remote Procedure Call, Message-Oriented/Stream-Oriented and Multicast; **Naming:** Names, Identifiers and Addresses, Flat, Structured and Attribute-Based; **Synchronization:** Physical Clocks, Global Positioning System, Clock Synchronization Algorithms, Logical Clocks, Lamport's Logical Clocks, Vector Clocks; **Mutual Exclusion:** A Centralized Algorithm, A Decentralized Algorithm, A Distributed Algorithm, A Token Ring Algorithm; **Election Algorithms:** Traditional Election Algorithms, Elections in Wireless Environments, Elections in Large-Scale Systems; **Consistency and Replication:** Data-Centric/ClientCentric Consistency Models; **Fault Tolerance:** Reliable Client-Server and Group Communication, Distributed Commit, Recovery; **Security:** Security Threats and Policies, Secure Channels, Security Management; **Distributed File Systems:** Architecture, Processes, Communication, Naming, Consistency and Replication, Fault Tolerance and Security, Sun NFS and Google FS (Case Studies).

References:

1. Steen, M V and Tanenbaum, A S, *Distributed Systems: Principles and Paradigms*, (2e), Pearson Education India, ISBN-10: 933254980X, ISBN-13: 978-9332549807, 2015.
2. Burns, B, *Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable Services*, (1e), ISBN-10: 1491983647, ISBN-13: 978-1491983645, 2018.
3. Ghosh, S, *Distributed Systems: An Algorithmic Approach*, (2e), Chapman and Hall / CRC Computer and Information Science Series, ISBN-10: 1466552972, ISBN-13: 978-1466552975, 2014.
4. Sinha, P K, *Distributed Operating Systems: Concepts and Design*, (1e), Prentice Hall India, ISBN10: 9788120313804, ISBN-13: 978-8120313804, 1998.

IT6141: DATA PREPARATION AND ANALYSIS [3 0 0 3]

Data Gathering and Preparation: Data Formats, Parsing and Transformation, Scalability and Real-time Issues; **Data Cleaning:** Consistency Checking, Heterogeneous and Missing Data, Data Transformation and Segmentation; **Exploratory Analysis:** Descriptive and comparative statistics, Clustering and association, Hypothesis Generation; **Visualization:** Designing Visualizations, Time Series, Geolocated Data, Correlations and Connections, Hierarchies and Networks, Interactivity.

References:

1. Glenn J. Myatt., *Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining*, (2e), Wiley Press, 2006.
2. E. Tufte. *The Visual Display of Quantitative Information*, (2e), Graphics Press, 2007.
3. Jules J., Berman D., *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*, (2e), 2013.

IT6142: STORAGE AREA NETWORK [3 0 0 3]

Storage Media and Technologies: Magnetic, Optical and Semiconductor Media, Techniques for read/write Operations, Issues and Limitations; **Usage and Access:** Positioning in the Memory Hierarchy, Hardware and Software Design for Access, Performance Issues; **Large Storages:** Hard Disks, Networked Attached Storage, Scalability Issues, Networking Issues; **Storage Architecture:** Storage Partitioning, Storage System Design, Caching, Legacy Systems; **Storage Area Networks:** Hardware and Software Components, Storage Clusters/Grids; **Storage QoS:** Performance, Reliability, and Security issues. Recent Trends Related to Copy Data Management, Erasure Coding, and Software-Defined Storage Appliances.

References:

1. Computer Technology Research Corporation, *The Complete Guide to Data Storage Technologies for Network-Centric Computing*, (1e), 1998.
2. Nigel Poulton.,*Data Storage Networking: Real World Skills for the CompTIA Storage+ Certification and Beyond*, (1e), Wiley Press, 2015.

IT6143: KNOWLEDGE DISCOVERY [3 0 0 3]

Introduction to Knowledge Management: Data, Information and Knowledge, Types of Knowledge, Knowledge Management Models, Organizational Memory, Knowledge Creation and Organizational Learning, Knowledge Codification, Knowledge Sharing and Transfer, Business and Competitive Intelligence, Ethical Issues in Knowledge Management, The Role of Culture in Knowledge Management, Business Process and Knowledge Management, Alignment of Business and Knowledge Management Strategies, Intellectual Capital and Knowledge Management, Measurement of Impact of Knowledge management programs; **Technologies for Knowledge Management:** Artificial Intelligence, Digital Libraries, Repositories, ECM, Knowledge-Based Systems, Information/Knowledge Audit; **An Architecture for Knowledge Discovery:** KM Cycle its Vision and Search, Generation, Acquisition, Capture, Transformation, Transfer, Application; **Knowledge Capture Systems:** Systems that Preserve and Formalize Knowledge; Concept Maps, Process Modeling, RSS, Wikis, Delphi Method; **Knowledge Sharing Systems:** Systems that Organize and Distribute Knowledge; Ontology Development Systems, Categorization and Classification Tools, XML-Based Tools; **Knowledge Discovery Methods:** Correlation, Class, Novelty, Association, Preprocessing Methods, Supervised Methods, Unsupervised Methods, Soft Computing Methods, Supporting Methods , Advanced Methods.

References:

1. Wesley, W Chu, *Data Mining and Knowledge Discovery for Big Data: Methodologies, Challenge and Opportunities*, (1e), Springer, 2013.
2. Dalkir, K, *Knowledge Management in Theory and Practice*, (2e), MIT Press, 2011.
3. Oded, M and Lior, R, *The Data Mining and Knowledge Discovery Handbook*, (2e), Springer, 2010.

IT 6243: RECOMMENDER SYSTEMS [3 0 0 3]

Introduction to Recommender Systems: Non-Personalized and Content-Based -- Introducing Recommender Systems, Non-Personalized and Stereotype-Based Recommenders, Content-Based Filtering; **Nearest Neighbor Collaborative Filtering**: User-User Collaborative Filtering Recommenders, Item-Item Collaborative Filtering Recommenders, Advanced Collaborative Filtering Topics; **Recommender Systems:** Evaluation and Metrics--Basic Prediction and Recommendation Metrics, Advanced Metrics and Offline Evaluation, Online Evaluation. Evaluation Design; **Matrix Factorization and Advanced Techniques**: Matrix Factorization, Hybrid Recommenders.

References:

1. K.Falk, *Practical Recommender Systems*, (1e), Manning Publication, 2017.
2. B.Chen, D.Agarwal, *Statistical Methods for Recommender Systems*", (1e), Cambridge University Press, 2015.

IT6244: WEB ANALYTICS AND DEVELOPMENT [3 0 0 3]

Introduction: Web Analytics, Web Analytics 1.0 vs. 2.0 Framework (Clickstream, Multiple Outcomes Analysis, Experimentation and Testing, Voice of Customer, Competitive Intelligence, Insights). **Working Of Web Analytics**: Basic Concepts, Basic Segmentation, Intermediate Metrics, Custom Metrics, and Calculated Metrics. **Data Types of Web Analytics**: Web Data and Other Types of Data, Also Basic Dashboards, Reports to Deliver Web Analytics. Web Analytics Ecosystem and Deploying It in Industry What To Measure. **Segmentation in Web Analytics:** Tracking, Visualization of Data, Acquisition and Conversions. Tracking Of Mobile Visitors, Other Web Analytics Reports and Visualizations. Third-Party Data and Comscore, Cohort Analysis and User Explorer. Geo-Social Data, Capstone Work, Web Analytics Case Studies.

References:

1. Avinash Kaushik, *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity*, (2e), John Wiley & Sons, 2018.
2. Tom Tullis, Bill Albert, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*, (2e), Morgan Kaufmann, July 2013.
3. Jim Sterne, *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*, (1e), John Wiley & Sons, 2010.
4. Brian Clifton, *Advanced Web Metrics with Google Analytics*, (3e), John Wiley & Sons, 2012.

IT6245: DATA ACCESS CONTROL AND SECURITY [3 0 0 3]

Introduction to Access Control: Purpose and Fundamentals of Access Control, Brief History, Policies of Access Control, Models of Access Control and Mechanisms, Discretionary Access Control (DAC), NonDiscretionary Access Control (NDAC), Mandatory Access Control (MAC); **Capabilities and Limitations of Access Control Mechanisms:** Access Control List (ACL) and Limitations, Capability List and Limitations; **Role-Based Access Control (RBAC) and Limitations**: Core RBAC, Hierarchical RBAC, Statically Constrained RBAC, Dynamically Constrained RBAC, Limitations of RBAC, Comparing RBAC to DAC and MAC Access Control Policy, Biba's Intrigity Model, Clark-Wilson Model, Domain Type Enforcement Model, Mapping the Enterprise View to the System View, Role Hierarchies, Inheritance Schemes, Hierarchy Structures and Inheritance Forms, SOD in Real System Temporal Constraints in RBAC, MAC and DAC; **Smart Card Based Information Security**: Smart Card Operating System-Fundamentals, Design and Implantation Principles, Memory Organization, Smart Card Files, File Management, Atomic Operation, Smart Card Data Transmission ATR, PPS Security Techniques, User Identification, Smart Card Security, Quality Assurance and Testing, Smart Card Life Cycle-5 Phases, Smart Card Terminals; **Recent Trends in Database Security and Access Control Mechanisms**: Case Study of RBAC Systems, Recent Trends Related to Data Security Management, Vulnerabilities in Different DBMS.

References:

1. David F, Ferraiolo D, Richard K, and Chandramouli R, *Role Based Access Contro*l, (1e), Artech House, ISBN: 1-58053-370-1, 2003.
2. Gerardus B*, Role-Based Access Control a Complete Guide,* (1e)*,* Emereo Pty Limited, 2019.
3. http://www.smartcard.co.uk/tutorials/sct-itsc.pdf : Smart Card Tutorial.