DEPARTMENT OF DATA SCIENCE AND COMPUTER APPLICATIONS

M.TECH (DATA SCIENCE)

Program Structure (Applicable to 2023 admission onwards)

YEAR	FIRST SEMESTER						SECOND SEMESTER					
	SUB CODE	**SUBJECT NAME**	**L**	**T**	**P**	**C**	**SUB CODE**	**SUBJECT NAME**	**L**	**T**	**P**	**C**
I	MAT 5129	MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE	3	1	0	4	DSE 5211	ARTIFICIAL INTELLIGENCE	3	1	0	4
	DSE 5111	DATA STRUCTURES AND ALGORITHMS	4	0	0	4	DSE 5212	DEEP LEARNING	3	1	0	4
	DSE 5112	DATABASE TECHNOLOGIES	3	0	0	3	DSE ****	PROGRAM ELECTIVE I	3	0	0	3
	DSE 5113	DATA ANALYTICS & VISUALIZATION	3	0	0	3	DSE ****	PROGRAM ELECTIVE II	3	0	0	3
	DSE 5114	MACHINE LEARNING	3	0	0	3	DSE ****	PROGRAM ELECTIVE III	3	1	0	4
	HUM 5051	RESEARCH METHODOLOGY & TECHNICAL COMMUNICATION*	1	0	3	-	*** ****	OPEN ELECTIVE	3	0	0	3
	DSE 5141	DATA STRUCTURES AND ALGORITHMS LAB**	0	0	3	2	HUM 5051	RESEARCH METHODOLOGY & TECHNICAL COMMUNICATION*	-	-	-	2
	DSE 5142	DATABASE TECHNOLOGIES LAB **	0	0	3	2	DSE 5241	AI AND DL LAB**	0	0	3	2
	DSE 5143	DATA ANALYTICS & VISUALIZATION LAB**	0	0	3	2	DSE 5242	DATA SCIENCE CAPSTONE PROJECT**	0	0	3	2
	TOTAL		**17**	**1**	**12**	**23**						**27**
	THIRD AND FOURTH SEMESTER											
II	DSE 6091	PROJECT WORK & INDUSTRIAL TRAINING							**0**	**0**	**0**	**25**

*TAUGHT IN BOTH SEMESTERS AND EVALUATED AND CREDITED IN THE SECOND SEMESTER

**LAB 3 Hrs/WEEK WITH A PROVISION FOR MINI PROJECT/ASSIGMENTS

PROGRAM ELECTIVES		OPEN ELECTIVES	
COURSE CODE	**COURSE TITLE**	**COURSE CODE**	**COURSE TITLE**
	PROGRAM ELECTIVE - I	DSE 5301	ENTERPRISE DATA ARCHITECTURE
DSE 5401	BIG DATA ANALYTICS	DSE 5302	BUSINESS INTELLIGENCE
DSE 5402	SOFT COMPUTING		
DSE 5403	COMPUTER NETWORK ANALYSIS		
ICT 5405	CLOUD AND EDGE COMPUTING		
ICT 5423	FEDERATED LEARNING		
	PROGRAM ELECTIVE - II		
DSE 5404	OPTIMIZATION TECHNIQUES FOR ANALYTICS		
DSE 5405	DATA PRIVACY AND SECURITY		
DSE 5406	PATTERN RECOGNITION & APPLICATIONS		
ICT 5402	SEMANTIC WEB TECHNOLOGIES		
ICT 5409	APPLIED NATURAL LANGUAGE PROCESSING		
	PROGRAM ELECTIVE - III		
DSE 5407	GRAPH ANALYSIS		
DSE 5408	COMPUTER VISION		
ICT 5414	TIME SERIES ANALYSIS AND FORECASTING		
ICT 5406	KNOWLEDGE AND DATA ENGINEERING		
ICT 5407	INFORMATION RETERIEVAL SYSTEMS		
ICT 5410	DECISION INTELLIGENCE		

DSE 5129 MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE [3 1 0 4]

Descriptive statistics-univariate, multivariate, sampling-estimation and inference. Stochastic Processes – types, fundamental concepts, limiting and transient behaviour. Birth-death models, Multivariate analysis-normal distribution and its transformed (Wishart)-properties & their uses in testing hypothesis, classification methods and multidimensional scaling.

SDL:
Graphical representation of multivariate data

References:

1. William J Stewart, *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*, Princeton University Press (2009).
2. Hardle, W.K & Simar, L, *Applied Multivariate Statistical Analysis*, Springer-Verlag Berlin Heidelberg (2014).
3. Dean W. Wichern and Richard A. Johnson, *Applied Multivariate Statistical Analysis*, Pearson (2007).
4. *Gupta, S.C and Kapoor, V.K. Fundamentals of Mathematical Statistics, Sultan Chand & Sons,11th Edition, 2016*
5. U. Narayan Bhat, *An Introduction to Queueing Theory-Modeling and Analysis in Applications*, Second Edition, 2015,. Birkhauser
6. George Casella and Roger L Berger, *Statistical Inference*, Second edition 2021,. Cengage Learning.

DSE 5111 DATA STRUCTURES AND ALGORITHMS [4 0 0 4]

Introduction, Data types, List, arrays, 2D arrays, Linked lists, Stacks, Queues, Asymptotic Analysis, Mathematical Analysis of Non-recursive and Recursive Algorithms, Trees-Binary trees, Binary Search Tree, Expression Tree, Threaded Binary Tree, Set, Heap, AVL tree, B-tree, B+ tree, Huffman tree, Hashing, Graphs, Topological Sort, Minimum Spanning Trees, Shortest Paths, Bipartite Graphs, Algorithm Design Techniques, Divide and Conquer, Greedy, Dynamic programming, Backtracking, Branch and Bound, Polynomial time, NP-completeness and reducibility.

SDL:
Polynomial time, Polynomial-time verification, NP-completeness and reducibility, NP-complete problems,

References:

1. Alfred V. Aho, John E. Hopcroft and Jeffrey D. Ullman, "*Data Structures and Algorithms*", 4th Edition, Addison Wesley, 2009.
2. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein "*Introduction to Algorithms*", 3rd Edition, PHI Publications, 2009.
3. Sartaj Sahni, "*Data Structures, Algorithms and Applications in C++",* 2nd Edition, Universities Press, 2005.
4. J. P. Trembley and Sorenson, *"An Introduction to Data Structures with Applications*" 2nd Edition, 36th Reprint, McGraw Hill, 2008.

DSE 5112 DATABASE TECHNOLOGIES [3 0 0 3]

Introduction Database Systems, Database Languages, Data models, Database architecture. Relational Model, keys, schema diagrams. Entity-Relationship Model, Reduction ER to Relational schema. Introduction to SQL Data Definition, constraints, importing and loading data, inserting, modification of data, Basic structure of SQL queries, Basic operations, Joins, Nested subqueries, Data Grouping, top-N-Queries, using regular expressions, views.PL/SQL, Cursors, Functions, Procedures, Triggers. Data Warehouse, Data Warehousing multi-tiered Architecture, Enterprise Warehouse, Data Mart, Schemas for Multidimensional Data Models, OLAP Operations, A StarNet Query Model for Querying Multidimensional Databases. Data Interchange formats, Understanding XML, constructing XML data from relational results sets, querying XML. Understanding JSON, format, JSON v/s XML, construct JSON data from relational results sets and querying.Unstructured data handling, the emergence of NoSQL, Aggregate model, Four Types of NoSQL Database, scheme less databases, materialized Distribution models-single server, Shading, Master-Slave replication, Peer-to-Peer Replication, combinations, relaxation consistency, Brewer's cap theorem,

SDL:
CRUD operations using MongoDB, Graph Databases, Features, Suitable Use Cases, CRUD operations using Neo4j/MongoDB.

References:

1. Abraham Silberschatz, Henry Korth, S. Sudarshan, *Database System Concepts,* 6th Edition, McGraw Hill, 2010.
2. Ramez Elmasri, Shamkant Navathe, *Fundamentals of Database System,* 6th Edition, Addison Wesley Publications Co., 2010.
3. Pramod J Sadalage, Martin Fowler, *NoSQL Distilled,* Addison-Wesley, 1st Edition, 2012.
4. Shashank Tiwari, *Professional NOSQL,* John Wiley & Sons Inc., 1st Edition, 2011.
5. Kyle Banker, Peter Bakkum, Shaun Yerch, Douglas Garrett, Tim Hawkins, *MongoDBin Action,* 2nd Edition, Manning Publications, 2016.
6. Kristina Chodorow, *MongoDB: The Definitive Guide*, O'Reilly publications,2nd Edition, 2013.
7. Peter A. Carter, *SQL Server Advanced Data Types JSON, XML, and beyond*, 2018.
8. Anders Moller and Michael I Schwartzbach, *Introduction to XML and Web Technologies,* Addison-Wesley, 2006.
9. Tom Marrs, *JSON at Work Practical Data Integration for the Web,* O'Reilly publications, 2017.

DSE 5113 DATA ANALYTICS & VISUALIZATION [3 0 0 3]

Steps in Data Analytics Projects, Data Analytics tasks, and methods, 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, Hypothesis testing, Statistical Inference. Association rule mining: Apriori, FP Growth, Partitioning, measures of pattern interestingness. Clustering: Partitioning,

Hierarchical, Density based approaches. Recommender Systems, Anomaly Detection. Visualization: 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, 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.

SDL:
Density based approaches. Recommender Systems, Anomaly Detection, Trends in Various Perception Techniques: Various Visualization Techniques, Data Structures used in Data Visualization

References:

1. Glenn J. Myatt, Wayne P. Johnson, *Making Sense of Data I: A Practical Guide to Exploratory Data Analysis and Data Mining*, 2nd Edition, John Wiley & Sons Publication, 2014.
2. Glenn J. Myatt, Wayne P. Johnson, *Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications*, John Wiley & Sons Publication, 2009.
3. Dr Anil Maheshwari*, "Data Analytics Made Accessible,* 2021 Edition.
4. Jules J., Berman D., *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*, (2e), 2013.
5. Matthew Ward and Georges Grinstein, *Interactive Data Visualization: Foundations, Techniques, and Applications*, (2e), A K Peters/CRC Press, 2015.
6. Jurgen Kai-Uwe Brock, *Data Design: The Visual Display of Qualitative and Quantitative Information*, (1e), Consulting Press, 2017.
7. Edward R. Tufte, *The Visual Display of Quantitative Information*, (2e), Graphics Press USA, 2001.
8. Cole Nussbaumer Knaflic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*, (1e), John Wiley and Sons, 2015.

DSE 5114 MACHINE LEARNING [3 0 0 3]

Machine Learning Basics: Types of Machine Learning, supervised vs. Unsupervised Learning, Parametric vs. non-parametric models., Instance Based learning – k-nearest neighbors, Simple Regression Models: Linear, Logistic, Cost functions, Gradient Descent, Batch Gradient Descent, Overfitting, Model Selection, No free lunch theorem, bias/variance trade-off, union and Chernoff bounds, VC dimensions. Bayesian Models: Bayesian concept learning, Bayesian Decision Theory, Naïve Bayesian, Laplacian Correction, Bayesian Belief Networks. Tree Models: information theory, decision tree induction, tuning tree size, ID3, C4.5, CHAID, Decision Stump. Support Vector Machines: kernel functions. Regression Models: Ridge and Lasso Regression, GLM and the exponential Family. Bagging algorithm, Random Forests, Grid search and randomized grid search, Partial dependence plots. Ensembling and Boosting Algorithms: Concept of weak learners, Adaptive Boosting, Extreme Gradient Boosting (XGBoost). Artificial Neural Networks: Perceptron, Back propagation,

Hopfield Network. Curse of Dimensionality: Factor Analysis, Principal Component Analysis (PCA), Difference between PCAs and Latent Factors,

SDL:
Curse of Dimensionality: Factor Analysis, Principal Component Analysis (PCA), Difference between PCAs and Latent Factors

References:

1. K. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.
2. G. James, D. Witten, T Hastie, R Tibshirani, An introduction to statistical learning with applications in R, Springer, 2013.
3. J. Han, M. Kamber, J. Pei, Data Mining concepts and techniques, (2e), Morgan Kaufmann-Elsevier, 2011.
4. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, (2e), Springer, 2009.
5. T. M. Mitchell, Machine Learning, (Indian Edition), MacGraw Hill, 2017.
6. C. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 2019

HUM 5051 RESEARCH METHODOLOGY & TECHNICAL COMMUNICATION [1 0 3 -]

Research Methodology: Basic concepts: Types of research, Significance of research, Research framework. Sources of data, Methods of data collection. Research formulation: Components, selection and formulation of a research problem, Objectives of formulation, and Criteria of a good research problem. Research hypothesis: Criterion for hypothesis construction, Nature of hypothesis, Characteristics and Types of hypotheses, Elements of research design, Introduction to various sampling methods Sources of data, Collection of data, Research reports, references styles, Effective Presentation techniques, Research Ethics.

SDL:
Research article preparation

References:

1. Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*. John Wiley & Sons.
2. Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2013). *Business research methods*. Cengage Learning.
3. Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.
4. Donald R Cooper & Pamela S Schindler, *Business Research Methods*, McGraw Hill International, 2018.

DSE 5141 DATA STRUCTURES AND ALGORITHMS LAB [0 0 3 2]

Searching Algorithms, Sorting Algorithms, Mapping of 2-D arrays to 1-D arrays, Stacks and Queues (Implementation using arrays), Evaluation of mathematical expressions, Conversion of mathematical expressions, Stacks and Queues (Implementation using linked lists), Binary

Search Tree, Representation of a directed graph, Advanced sorting algorithms, Knapsack problem, Assignment problem.

SDL:
None

References:

1. Behrouz A. Forouzan, Richard F. Gilberg, *A Structured Programming Approach Using C*, (3e), Cengage Learning India Pvt. Ltd, India, 2007
2. Ellis Horowitz, Sartaj Sahni, Susan Anderson and Freed, *Fundamentals of Data Structures in C*, (2e), Silicon Press, 2007
3. Richard F. Gilberg, Behrouz A. Forouzan, *Data structures, A Pseudocode Approach with C*, (2e), Cengage Learning India Pvt. Ltd, India, 2009
4. Tenenbaum Aaron M., Langsam Yedidyah, Augenstein Moshe J., *Data structures using C*, Pearson Prentice Hall of India Ltd., 2007
5. Debasis Samanta, *Classic Data Structures*, (2e), PHI Learning Pvt. Ltd., India, 2010

DSE 5142 DATABASE TECHNOLOGIES LAB [0 0 3 2]

Implementation/realization of database concepts such as SQL, Queries, Join, Views, Advanced SQL, PL/SQL, Cursors, Functions and Procedures, Triggers, Mini project based on the concepts learnt.

SDL:
None

References:

1. Abraham Silberschatz, Henry Korth, S. Sudarshan, *Database System Concepts,* 6th Edition, McGraw Hill, 2010.
2. Ramez Elmasri, Shamkant Navathe, *Fundamentals of Database System,* 6th Edition, Addison Wesley Publications Co., 2010.
3. Pramod J Sadalage, Martin Fowler, *NoSQL Distilled,* Addison-Wesley, 1st Edition, 2012.
4. Shashank Tiwari, *Professional NOSQL,* John Wiley & Sons Inc., 1st Edition, 2011.
5. Kyle Banker, Peter Bakkum, Shaun Yerch, Douglas Garrett, Tim Hawkins, *MongoDBin Action,* 2nd Edition, Manning Publications, 2016.
6. Kristina Chodorow, *MongoDB: The Definitive Guide*, O'Reilly publications,2nd Edition, 2013.
7. Peter A. Carter, *SQL Server Advanced Data Types JSON, XML, and beyond*, 2018.
8. Anders Moller and Michael I Schwartzbach, *Introduction to XML and Web Technologies*, Addison-Wesley, 2006.
9. Tom Marrs, *JSON at Work Practical Data Integration for the Web*, O'Reilly publications, 2017

DSE 5143 DATA ANALYTICS & VISUALIZATION LAB [0 0 3 2]

Tutorial on tools for Data Analytics & Visualization. Suggested tools are Python, MATLAB, WEKA, R Studio. Experiments with datasets to be defined in lab manual, to implement

concepts of data preprocessing, exploratory analysis, comparative statistics, statistical inference, Association and clustering. Creating Visual Representations- Suggested tools are MSExcel, Power BI, Tableau.

SDL:
None

References:

1. Manaranjan Pradhan, U Dinesh Kumar, *Machine Learning using Python*, Wiley Publications.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer Science, 2017
3. Cole Nussbaumer Knaflic, *Storytelling with Data: A Data Visualization Guide for Business Professionals*, (1e), John Wiley and Sons, 2015.
4. Brian R. Hunt, Ronald L. Lipsman, Jonathan M. Rosenberg, Kevin R. Coombes, John E. Osborn, Garrett J. Stuck , *Guide to MATLAB: For Beginners and Experienced Users*,(2e), Cambridge University Press, 2011.

SECOND SEMESTER

DSE 5211 ARTIFICIAL INTELLIGENCE [3 1 0 4]

History of AI, Foundations of AI, AI Approaches - Cognitive Modeling. The Turing Test, Rational thinking – Logic. Intelligent Agents and Environments- the concept of Rationality, Single and Multi-Agents, Performance Evaluation of Agents, and Architecture of Intelligent Agents. AI Problems: Problem Space, Problem analysis. Problem-Solving Techniques; Heuristic Search, Uninformed Search, Local Search, Constraint Satisfaction Problems. Games: Optimal decision in games, Adversarial Search, Minimax, Alpha Beta Pruning, The Wumpus World. Reinforcement learning: Multi-armed bandits, Markov Decision Process, Bellman equations, Dynamic Programming, Monte Carlo methods, Temporal Difference Learning, Policy Gradient Methods. Knowledge and Reasoning, Representation: Logical Agents, First order logic and inference, Classical Planning. Propositional logic: Propositional Theorem Proving, Representation, Fuzzy Logic. Ontological Engineering, Semantic Technologies, RuleML. Quantifying Uncertainty, Probabilistic Reasoning, Making Simple & Complex Decisions. Applications: NLP, Parsing, Machine translation, speech recognition, Perception: Image formation, Image Processing, Object Recognition, Robotics: software agents, Hardware, perception, software architectures. Future of AI : Cognitive Modeling approach, Layers of Mental Activities, Layered Knowledge Representation, Cognitive Architectures, Brain –Machine Convergence.

SDL:
Case Studies in Quantifying Uncertainty, Probabilistic Reasoning, Making Simple & Complex Decisions.

References:

1. Stuart Russell and Peter Norvig, *Artificial Intelligence A Modern Approach* , Pearson 2010.
2. Marvin Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster, 2007.
3. Richard S Sutton and Andrew G Barto, *Reinforcement Learning: An Introduction*, MIT Press
4. Rich E., Knight K., Nair S.B., *Artificial Intelligence*, Tata McGraw Hill, 2008.
5. Grigoris Antoniou, Paul Groth, Frank van Harmelen and Rinke Hoekstra, *A Semantic Web Primer*, MIT Press, 2012.

DSE 5212 DEEP LEARNING [4 0 0 4]

Introduction, Neural Network Basics: Multi-layer perceptron, Back propagation algorithm, training procedures, Shallow Neural Networks: Review, Gradient descent, and Activation Function Deep Feed Forward Networks: Forward and Backward Propagation, Hidden units, architecture design, Dimensionality reduction, learning time. Regularization for Deep Learning: Parameter Norm Penalties, Regularization and Under-Constrained Problems, Dataset Augmentation, Noise-Robustness, Bagging and Other Ensemble Methods, Dropout, Adversarial Training. Optimization for Training Deep Models: Challenges in Neural Network Optimization. Deep Neural Networks and the Brain. Convolutional Networks: convolution operation, pooling Object detection and Face recognition Sequence Modelling: Recurrent and Recursive Networks, Stacked Auto Encoders: Under complete, Regularized, sparse, de-noising, Monte Carlo Methods. Markov Models, Hidden Markov models: evaluation problem, finding the state sequence, HMM as graphical model. Deep Generative Models Boltzmann Machines-the physics, randomness, impact on cognitive learning. Deep Boltzmann Machines, Deep Belief Networks-its relationship to Boltzmann Machines, concept of greedy networks, application to drug discovery. Generative Adversarial Networks, Auto-regressive Networks. Practical Methodology: Performance Metrics, Default Baseline Models, selecting hyper parameters, Debugging Strategies. Case Studies in: Large Scale Deep Learning, Computer Vision, Speech Recognition, Economics, Fraud detection, Crime detection.

SDL:
Case Studies in: Large Scale Deep Learning, Computer Vision, Speech Recognition, Economics, Fraud detection, Crime detection.

References:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning*, MIT Press 2016.
2. Alpaydin Ethem, *Introduction to Machine Learning*, 3rd Edition, PHI Learning Private Limited, 2018.
3. Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
4. Simon Haykin, Neural Networks and Learning Machines, PHI, 2008
5. Rajasekaran S., and Pai G. A. V., *Neural Networks, Fuzzy Logic and Genetic Algorithms*, PHI Learning, 2010.

PROGRAM ELECTIVE I

PROGRAM ELECTIVE II

PROGRAM ELECTIVE III

OPEN ELECTIVE

DSE 5241 AI AND DL LAB [0 0 3 2]

Tutorial on tools for Machine Learning. TensorFlow, Python-Keras suggested. Experiments with datasets to be defined in lab manual to deploy deep learning algorithms. Case studies or mini projects in topics such as Sentiment Analysis, Anomaly Detection, Recommender Systems.

SDL:
None.

References:

1. Hans Peter Langtangen, Python Scripting for Computational Science, (3e), Springer Publishers, 2014
2. Naomi R. Ceder, The Quick Python Book, (2e), Manning Publications Co., 2010
3. Ahmed Menshawy, Md. Rezaul Karim, Giancarlo Zaccone , Deep Learning with TensorFlow, Packet Publishing

DSE 5242 DATA SCIENCE CAPSTONE PROJECT [0 0 3 2]

A case study based on the standard public dataset/real time data and application of the technologieslearnt through all the courses which leads to a research publication.

SDL:
None.

References:
None

THIRD AND FOURTH SEMESTER

DSE 6091 PROJECT WORK & INDUSTRIAL TRAINING

PROGRAM ELECTIVE I

DSE 5401 BIG DATA ANALYTICS [3 0 0 3]

Introduction to Big Data: Evolution, Structuring, Elements, Big Data Analytics, Distributed and Parallel Computing for Big Data, Hadoop, Cloud Computing and Big Data, In-memory Computing Technology for Big Data, Big Data Stack, Virtualization and Big Data, Hadoop: Ecosystem, Hadoop Distributed File System (HDFS), MapReduce: MapReduce Framework, Optimizing MapReduce Jobs, MapReduce Applications, Understanding YARN architecture, HBase, Exploring Hive, Analyzing Data with Pig, Using Oozie, Introduction to Mahout, Role of HBase in Big Data Processing, RHadoop: Data Analysis Using the MapReduce Technique in RHadoop, Spark: Core Concepts, Spark's Python and Scala shells, Programming with RDD: RDD Operations, Passing Functions to Spark, Common Transformations and Actions, Mining Data Streams: Streams Concepts, stream Data Model and Architecture, Stream Computing, Filtering Streams, Estimating Moments, Decaying Window, Real time Analytics Platform (RTAP) Applications, Case studies: Real Time Sentiment Analysis, Stock Market Predictions.

SDL:
Data Streams: Streams Concepts, stream Data Model and Architecture, Stream Computing, Filtering Streams, Estimating Moments, Decaying Window, Real time Analytics Platform (RTAP) Applications

References:

1. *Big Data Black Book: (Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization).* Dreamtech Press, 2016.
2. Vignesh Prajapathi, *Big Data Analytics with R and Hadoop,* Packt Publishing, 2013.
3. Holden Karau, Andy Konwinski,Patrick Wendell,Matei Zaharia, *Learning Spark: Lightning-Fast Big Data Analysis*,1st Edition, O'Reilley Media Inc, 2015.
4. Michael Minnelli, Michele Chambers, *Big Data Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, Wiley India Pvt. Ltd., 2013.
5. Arvind Sathi, *Big Data Analytics*, MC Press, LLC, 2012.

DSE 5403 COMPUTER NETWORK ANALYSIS [3 0 0 3]

Introduction to Computer Networks, Network architectures, Types of networks, Network protocols, OSI and TCP/IP models, Network Traffic Analysis, Packet capture and analysis Network traffic flows, Analyzing network traffic patterns, Network traffic classification. Network Performance Evaluation, Metrics for network performance, Bandwidth, throughput, and latency, Queuing theory, Network simulation, Network Analysis Tools, Wireshark, Tcpdump, Nmap, Ping and Traceroute, Netcat, Applications of Network Analysis, Network troubleshooting and optimization, Quality of service (QoS), Network monitoring and management, Network forensic analysis.

SDL:
Applications of Network Analysis, Network troubleshooting and optimization, Quality of service (QoS), Network monitoring and management, Network forensic analysis

References:

1. Behrouz A. Forouzan, *TCP/IP Protocol Suite (3e),* Tata McGraw Hill,2012.
2. Leon-Garcia and Widjaja, *Communication Networks*, Tata McGraw Hill, 2010.
3. Jean Walrand and Pravin Varayya, *High Performance Communication Networks,* Harcourt Asia Pvt Ltd., 2010.
4. Andrew S. Tanenbaum, *Computer Networks (4e),* Prentice Hall India, 2012.
5. E. Van Valkenburg, "*Network Analysis*", Prentice Hall of India
6. A C.L Wadhwa*, "Network Analysis and Synthesis*" New Age International Publishers, 2007,
7. Roy Choudhary, *"Networks and Systems*" Wiley Eastern Ltd.
8. E. Van Valkenburg, "*An Introduction to Modern Network Synthesis*",Wiley Eastern Ltd.
9. Chakrabarti, *"Circuit Theory*" Dhanpat Rai & Co.

ICT 5423 SOFT COMPUTING [3 0 0 3]

Introduction to Soft Computing: Concept of Computing Systems, Soft Computing Versus Hard Computing, Characteristics of Soft Computing, Some Applications of Soft Computing Techniques; Fuzzy Logic: Introduction to Fuzzy Logic- Fuzzy Sets and Membership Functions, Operations on Fuzzy Sets, Fuzzy Relations, Rules, Propositions, Implications and Inferences, Defuzzification Techniques - Fuzzy Logic Controller Design, Some Applications of Fuzzy Logic; Artificial Neural Networks: Biological Neurons and its Working, Simulation of Biological Neurons to Problem Solving, Different ANNs Architectures, Training Techniques for ANNs, Applications of ANNs to Solve Some Real Life Problems; Nature Inspired Algorithms: Genetic Algorithms, Concept of "Genetics" and "Evolution" and its Application to Probabilistic Search Techniques, Basic GA Framework and Different GA Architectures, GA Operators, Solving Single-Objective Optimization Problems Using GAs, Particle Swarm Optimization, Case Studies, Ant Bee Colony Optimization, Case Studies; Multi-Objective Optimization: Problem Solving Concept of Multi-Objective Optimization Problems (MOOPs) and Issues of Solving Them. Multi-Objective Evolutionary Algorithm (MOEA), Non-Pareto Approaches to Solve MOOPs, Pareto-Based Approaches to Solve MOOPs, Some Applications with MOEAs.

SDL:
Multi-Objective Evolutionary Algorithm (MOEA), Non-Pareto Approaches to Solve MOOPs, Pareto-Based Approaches to Solve MOOPs, Some Applications with MOEAs

References:

1. Martin, F., Neill, Mc. and Thro, E., *Fuzzy Logic: A Practical approach*, AP Professional, 2000.
2. Ross, T, J., *Fuzzy Logic with Engineering Applications*, (3e), Willey India, 2010.

3. Kasabov, N. K., *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, MIT Press, 1998.
4. Ibrahim, A. M., *Fuzzy Logic for Embedded Systems Applications*, Elsevier Press, 2004.
5. Mitchell, M., *An Introduction to Genetic Algorithms*, MIT Press, 2000.
6. Goldberg, D. E., *Genetic Algorithms In Search, Optimization And Machine Learning*, Pearson Education India, 2002.
7. Rajasekaran, S. and Vijayalakshmi Pai, G. A., *Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis, and Applications*, Prentice Hall India, 2007.
8. Pratihar, D. K., *Soft Computing*, (1e), Narosa Publishing, 2008.
9. Jang, J.-S. R., Sun, C.-T. and Mizutani, E., *Neuro-Fuzzy and Soft Computing*", (1e), PHI Learning, 2009.
10. Haykin, S., *Neural Networks and Learning Machines*, (3e), Prentice Hall India, 2011

PROGRAM ELECTIVE II

DSE 5404 DATA PRIVACY AND SECURITY [3 0 0 3]

Introduction to Data Privacy, types of privacy attacks, Data linking and profiling, access control models, role-based access control, privacy policies, their specifications, privacy policy languages, privacy in different domains-medical, financial, etc. Mathematical model for comparing real-world data sharing practices, computing privacy and risk measurements. Demographics and Uniqueness. Protection Models-Null-map, k-map, Wrong map. Survey of Techniques-Protection models (null-map, k-map, wrong map), Disclosure control, Inferring entity identities, entry specific databases. Computation systems for protecting delimited Data-Min Gen, Datafly, Mu-Argus, k-Similar. Introduction to Security: The OSI Security Architecture, Security Attacks, Services and Mechanisms, Model for Network Security, Introduction to cryptography, Symmetric Encryption, Substitution ciphers, Stream ciphers, Data Encryption Standard, Number theory, Advance Encryption Standard Public-key cryptography, Digital Signatures, Cryptographic Hash Functions, Message Authentication and Confidentiality, Key Distribution and Authentication, Transport Layer Security, Wireless Network Security, E-mail Security, IP Security, Security Management Systems, Intrusion Prevention and Detection Systems. Security protocols, Secure Socket Layer (SSL), Intruders and Viruses, Firewalls

SDL:
Transport Layer Security, Wireless Network Security, E-mail Security, IP Security, Security Management Systems, Intrusion Prevention and Detection Systems. Security protocols, Secure Socket Layer (SSL), Intruders and Viruses, Firewalls

References:

1. Ronald Leenes , Rosamunde van Brakel , Serge Gutwirth , De Hert, Paul, *Data Protection andPrivacy: The Age of Intelligent Machines (Computers, Privacy*

and Data Protection), Hart Publishing, December 28, 2017.
2. B. Raghunathan*, The Complete Book of Data Anonymization: From Planning to Implementation,* Auerbach Pub, 2016.
3. L. Sweeney, *Computational Disclosure Control: A Primer on Data Privacy Protection,* MITComputer Science, 2017
4. William Stallings, *Cryptography and Network Security: Principles and Practice*, 7th Edition,Pearson Education, 2017.
5. William Stallings, *Network Security Essentials: Applications and Standards*, 6th Edition, PearsonEducation, 2014.
6. B. A. Forouzan, *Cryptography & Network Security,* Tata Mc Graw Hill.

DSE 5405 OPTIMIZATION TECHNIQUES FOR ANALYTICS [3 0 0 3]

Linear Programming: Problem Formulation, Linear Programming (LP) in standard form, Graphical Solution, Simplex Method, Big M Method. Transportation and Assignment Model: Transportation problem formulation, optimal solution, unbalanced transportation problem, degeneracy, Assignment problem, Hungarian Problem. Network Analysis: Graphs, Network and Flows, Minimum cost flow Models, Sources, Sinks Max Flow - Min Cut Theorem, CPM and PERT Networks. Inventory: Introduction, Single Item, Deterministic model, Purchase Inventory model with one price, break and multiple price breaks. Dynamic Programming: Introduction, Forward and backward recursions, Bellman's Principle of Optimality. Equipment Replacement Model, Allocation Problem, Inventory Models. Decision Theory: Decision under certainty: Analytic Hierarchy Process (AHP), decision under risk: decision trees, expected value criterion, Variations of the Expected value criterion, decision under uncertainty: Laplace, MinMax, Savage, Hurwicz method. Game Theory: Introduction, Minmax – Maxmin pure strategies, Optimal solution of two person zero sum games, solution of mixed strategy games, 2 x 2 games, 2 x n games, m x 2 games. Heuristics and approximation algorithms: approximation algorithms for Travelling Salesman Problem (TSP), Vertex cover problem.

SDL:

Heuristics and approximation algorithms: approximation algorithms for Travelling Salesman Problem (TSP), Vertex cover problem

References:

1. Taha H, *Operation Research: An Introduction*, 10th Edition, McMillan, 2017.
2. Rardin, Ronald L., *Optimization in Operations Research*, Pearson Education (2005)
3. Ravindra K. Ahuja, Thomas L. Magnanti, James B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Pearson New International Edition, 2014.
4. Teofilo F. Gonzalez, *Handbook of Approximation Algorithms and Metaheuristics*, Chapman & Hall/CRC Computer and Information Science Series, 1st Edition, 2007.
5. S.S. Rao, *Engineering Optimization: Theory and Practice*, New Age International Pvt. Ltd., New Delhi, 2013.

DSE 5406 PATTERN RECOGNITION & APPLICATIONS [3 0 0 3]

Introduction: Definitions of data sets for Pattern Recognition (PR), Different paradigms of PR, Representations of Patterns and Classes, Metric and Non-metric proximity measures, Applications of PR, Feature extraction and feature selection: Feature extraction, different approaches to feature selection, Feature ranking. Statistical Decision Making: Introduction, Bayes theorem, multiple features, conditionally independent features, decision boundaries, the leaving-one-out technique, characteristic curves, estimating the composition of populations. Naïve Bayes classifier, Bayesian Belief Networks, Supervised and unsupervised Classification: Introduction to supervised and unsupervised classifications, Classification in High dimension, Random forests, SVM classifications. Introduction to clustering, clustering large datasets and combination of classifiers.

SDL:
Classification in High dimension, Random forests, SVM classifications. Introduction to clustering, clustering large datasets and combination of classifiers.

References:

1. Devi V. S, Murthy M. N, *Pattern Recognition: An Introduction,* Universities Press, Hyderbad. 2011.
2. Earl Gose, Richard Johnsonbaugh and Steve Jost, *Pattern Recognition and Image Analysis*, Prentice Hall of India, 2003.
3. R.D. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, 2nd Edition, John Wiley Inc., 2001.

PROGRAM ELECTIVE III

DSE 5407 COMPUTER VISION [3 1 0 4]

Introduction : Image Processing, Components of Image processing system, Image formation and digitization concepts, Neighbors of pixel adjacency connectivity, regions and boundaries, Distance measures, Image processing operations, Arithmetic, Logical, Geometrical, Convolution and Correlation Operations, Applications: Document Image Analysis, Biometrics, Object Recognition, Tracking, Medical Image Analysis, Content-Based Image Retrieval, Video Data Processing, Multimedia, Virtual Reality and Augmented Reality, Image Formation, Image representations (continuous and discrete) , Image pre-processing Techniques, Feature Extraction-Point, Line and Edge Detection, Color, Texture, Shape and structure Features in spatial and frequency domains, Corner Detection, Hough Transform , Image Segmentation: Boundary detection based techniques, Point, line detection, Edge detection, Edge linking, local processing, regional processing, Hough transform, Thresholding, Iterative thresholding, Otsu's method, Moving averages, Multivariable thresholding, Region-based segmentation, Watershed algorithm, Use of motion in segmentation, Computer Vision: Computer Vision, What is Computer Vision - Low-level, Mid-level, High-level , Overview of Diverse Computer Vision Applications,Fundamentals of object recognition, Low-level computer vision-Edges, contours, textures, shapes, and colors , Motion, optical flow, and tracking Local features,

invariance, bag-of-words models, Fisher vector, Middle-level representations of objects: parts, attributes, embedding.

SDL:

Motion, optical flow, and tracking Local features, invariance, bag-of-words models, Fisher vector, Middle-level representations of objects: parts, attributes, embedding.

References:

1. David A forsyth & Jean ponce, *Computer vision – A modern Approach*, Prentice Hall, PearsonEducation India; Edition: Second, 2015.
2. R. C. Gonzalez, R. E. Woods. *Digital Image Processing*, Pearson, Inc., Edition-Fourth.
3. K. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall, Pearson; Edition: First
4. Bernd Jahne and Horst HauBecker, *Computer vision and Applications*, Academic Press, 2000.

DSE 5408 GRAPH ANALYSIS [3 1 0 4]

Introduction to Graphs, Types of graphs, Basic graph terminology, Graph representations, Network Analysis, Network properties, Degree distribution, Centrality measures, Clustering coefficient, Small-world networks, Scale-free networks, Graph Algorithms, Breadth-first search, Depth-first search, Shortest path algorithms, Minimum spanning trees, Maximum flow algorithms, Visualization, Basic visualization techniques, Layout algorithms, Dynamic visualization, Interactive visualization, Applications of Graph Analysis, Social network analysis, Web analysis, Biological network analysis, Transportation network analysis, Tools and Libraries for Graph Analysis, NetworkX, Gephi, Other tools and libraries

SDL:

Transportation network analysis, Tools and Libraries for Graph Analysis, NetworkX, Gephi, Other tools and libraries

References:

1. J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*, Second. Cambridge University Press, 2014.
2. S. Even, *Graph Algorithms*, 2nd ed. Cambridge: Cambridge University Press, 2011.
3. Easley David and Kleinberg Jon. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, USA., 2010
4. Ian Robinson, Jim Webber, Emil Eifrem, *Graph Databases*, 2nd Edition, O'Reilly Media, Inc., 2015