M. Tech.

Computer Science and Engineering

(CSE)

with Specialization in

Data Science

At the end of studying the program, a student is expected to

1. engage in critical thinking and develop an ability to independently carry out research /investigation and development work to solve practical problems.
2. develop an ability to communicate effectively, develop an ability to interact with the engineering fraternity and with society at large.
3. be able to write and present technical reports on complex engineering activities.
4. be able to demonstrate a degree of mastery over the area as per the specialization of the program (Data Science). The mastery should be at a level higher than the requirements in the appropriate bachelor program.
5. demonstrate higher level of professional skills to tackle multidisciplinary and complex problems related to variety real time applications data.
6. be able to distinguish and analyze the data for the applications for the machine-cognition tasks.
7. have adequate technologies and theoretical background of software development that will help them to pursue a career in software industries in general and data science background in particular.
8. be educated to stick on professional ethics and able to solve societal needs and developments.

M. Tech. - I Computer Science and Engineering (CSE) with Specialization in Data Science

Semester I

Sr. No.	Course	Code	Credit	Teaching Scheme			Examination Scheme			Total
				L	T	P	L	T	P	
1.	Core-1 Mathematical Foundations of Computer Science	CSEDS601	4	3	1	0	100	25	0	125
2.	Core-2 Design and Analysis of Algorithms	CSEDS603	4	3	0	2	100	0	50	150
3.	Core-3 Machine Learning	CSEDS605	4	3	0	2	100	0	50	150
4.	Core-4 Foundations of Data Science	CSEDS607	4	3	0	2	100	0	50	150
5.	Core Elective-1	CSEDSXXX	4	3	0	2	100	0	50	150
6.	Research Methodology in CSE	CSEDS609	4	4	0	0	100	0	0	100
	Total		**24**	**19**	**1**	**8**	**600**	**25**	**200**	**825**
	Total Contact Hours per week			**28**						

Semester II

Sr. No.	Course	Code	Credit	Teaching Scheme			Examination Scheme			Total
				L	T	P	L	T	P	
1.	Core-5 Advanced Statistical Techniques	CSEDS602	4	3	1	0	100	25	0	125
2.	Core-6 Scalable Systems for Data Science	CSEDS604	4	3	0	2	100	0	50	150
3.	Core Elective-2	CSEDSXXX	4	3	0	2	100	0	50	150
4.	Core Elective-3	CSEDSXXX	4	3	0	2	100	0	50	150
5.	Core Elective-4	CSEDSXXX	4	3	0	2	100	0	50	150
6.	Institute Elective	CSEDSXXX	4	3	0	2	100	0	50	150
	Total		**24**	**18**	**1**	**10**	**600**	**25**	**250**	**875**
	Total Contact Hours per week			**29**						

Semester III

Sr. No.	Course	Code	Credit	Teaching Scheme			Examination Scheme			Total
				L	T	P	L	T	P	
1.	MOOC-I*	CSEDS701	2	2	0	0	50	0	0	50
2.	MOOC-II*	CSEDS703	2	2	0	0	50	0	0	50
3.	Dissertation Preliminaries#	CSEDS705	8	0	0	16	0	0	250	250
	Total		**12**	**4**	**0**	**16**	**100**	**0**	**250**	**350**
	Total Contact Hours per week			**20**						

***NPTEL, SWAYAM and other Massive Open Online Course (MOOC) approved by DAAC**

 # Internal-100, External-150

Semester IV

Sr. No.	Course	Code	Credit	Teaching Scheme			Examination Scheme			Total
				L	T	P	L	T	P	
1.	Dissertation#	CSEDS700	12	0	0	24	0	0	400	400
	Total		**12**	**0**	**0**	**24**	**0**	**0**	**400**	**400**
	Total Contact Hours per week			**24**						

 # Internal-160, External-240

Core Elective 1	
CSEDS611	Information Retrieval
CSEDS613	Advanced Database Management Systems
CSEDS615	Embedded Systems Design
CSEDS617	Computer Vision and Image Processing
CSEDS619	Speech and Audio Processing
CSEDS621	High Performance Computing
Core Elective 2, Core Elective 3, and Core Elective 4	
CSEDS606	Artificial Intelligence
CSEDS608	Data Mining and Data Warehousing
CSEDS610	Natural Language Processing
CSEDS612	Data Science for Software Engineering
CSEDS614	Big Data Analytics and Large-Scale Computing
CSEDS616	Cyber Physical Systems
CSEDS618	Machine Learning for Security
Institute Elective	
CSEDS620	Business Data Analytics
CSEDS622	Social Networks
CSEDS624	Cyber Laws

M. Tech. – I Semester – I	L	T	P	C
CSEDS601: MATHEMATICAL FOUNDATIONS OF COMPUTER SCIENCE (CORE-1)	**3**	**1**	**0**	**4**

Course Objective	
1	To learn the fundamental concepts of set theory, functions, probability.
2	To enable the students to apply the knowledge of probability in data science applications.
3	To learn different statistical inference procedures, probability distributions and random processes.
4	To enable the student to apply the knowledge of linear algebra and statistical analysis in different fields of data science.
5	To design an efficient solution using linear algebra and statistical methods for real time problems.

INTRODUCTION	**(06 Hours)**
Set Theory, Logic and Proofs, Conditional Propositions, Logical Equivalence, Predicates, Quantifiers, Combinatorics.	
FUNCTIONS AND RELATIONS	**(06 Hours)**
Types of Functions, Recursive Functions, Computable and non-computable Functions, Representations of Relations, Composition and Properties of Relations.	
PROBABILITY AND RANDOM VARIABLES	**(10 Hours)**
Overview of Sample Points and Sample Spaces, Events, Bayes Theorem, Probability Axioms, Joint and Conditional Probability, Random Variables, Discrete and Continuous Random Variables, Random Vectors, Transformation of Continuous Random Variables and Vectors by Deterministic Functions, Density Functions of Transformed Continuous Random Variables and Vectors, Multivariate Random Variables, Moments and Moment Generating Functions, Functions of Random Variables.	
RANDOM PROCESSES	**(10 Hours)**
Random Variable vs. Random Process, Bernoulli Random Process, Binomial Process, Statistical Averages, Ensemble and Time Averages, Weak and Strict Sense Stationarity of a Random Process, Ergodicity, Autocorrelation and Auto Covariance Functions of Random Processes and its Relation to Spectra, Poisson Process, Gaussian Process, Martingale Model and Markov Chains.	
ESTIMATION AND STATISTICAL ANALYSIS	**(10 Hours)**
Estimation of Parameters from Data, Maximum Likelihood Estimation, Maximum a Posterior Estimation, Consistency and Efficiency of Estimators, Stochastic State Estimation and MSE of an Estimator, Estimation of Gaussian Random Vectors, Linear Minimum Mean Square Error Estimation, Hypothesis Testing, Significance	

Level, Types of Errors: Type-I and Type-II, Significance Test, Chi-Squared, Student-t Test, Normality Test, Cramer-Rao Bound on Estimators, Chebyshev Inequality, Kullback-Leibler Divergence, Applications.	
Tutorial Assignments will be based on the coverage of above topics. (Problem statements will be changed every year and will Be notified on website.)	**(14 Hours)**
	(Total Contact Time: 42 Hours + 14 Hours = 56 Hours)

BOOKS RECOMMENDED (LATEST EDITION)

1. Keneth H. Rosen, "Discrete Mathematics and Its Applications", McGraw-Hill.
2. Judith L. Gersting, "Mathematical Structure for Computer Science", W.H. Freeman and Co.
3. Athanasios Papoulis and S. Unnikrishna Pillai, "Probability, Random Variables and Stochastic Processes", McGraw-Hill.
4. Wilbur B. Davenport, "Probability and Random Processes - an introduction for application scientists and engineers", McGraw-Hill.
5. Sheldon M. Ross, Introduction to Probability Models", Academic Press.

Course Outcomes **At the end of the course, students will**	
CO1	have knowledge of the basic concepts and problems of set theory, predicates and logic.
CO2	be able to use functions, graphs, trees, automata and formal languages for problem solving.
CO3	be able to analyze/interpret quantitative data verbally, graphically, symbolically and numerically.
CO4	be able to evaluate and compare the results using different linear algebraic and statistical techniques.
CO5	be able to use linear algebra for optimization and integrate statistical models for solving real life applications.

M. Tech. – I Semester – I	L	T	P	C
CSEDS603: DESIGN AND ANALYSIS OF ALGORITHMS (CORE-2)	**3**	**0**	**2**	**4**

Course Objective	
1	To understand paradigms and approaches used to analyze and design algorithms and to appreciate the impact of algorithm design in practice.
2	To analyze the worst-case time complexity of an algorithm, asymptotic complexities of different algorithms.
3	To design and prove the correctness of the algorithms using appropriate design technique to solve a given real-world computational problem.
4	To analyze and prove the computational intractability of the algorithms of the hard computational problems.
5	To design sub-optimal solutions for the intractable computational problems using alternate design approaches.

INTRODUCTION	**(02 Hours)**
Review of Basis Concepts in Algorithms, Abstract Machines, Analysis Techniques: Mathematical, Empirical and Asymptotic analysis, Review of the Notations in Asymptotic Analysis, Recurrence Relations and Solving Recurrences, Proof Techniques, Illustrations.	
DIVIDE AND CONQUER APPROACH	**(06 Hours)**
Review of Sorting & Order Statistics, Various Comparison based Sorts Analysis, Medians and Order Statistics, The Union-Find Problem, Counting Inversions, Finding the Closest Pair of Points; Lower Bound on Sorting and Non-comparison based Sorts.	
SEARCHING AN DSET MANIPULATION	**(02 Hours)**
Searching in Static Table Binary Search, Path Lengths in Binary Trees and Applications; Optimality of Binary Search in Worst Case and Average Case; Binary Search Trees, Construction of Optimal Weighted Binary Search Trees; Searching in Dynamic Table, Randomly Grown Binary Search Trees, AVL and (a, b) Trees.	
HASHING	**(02 Hours)**
Basic Ingredients, Analysis of Hashing with Chaining and with Open Addressing; Union-Find Problem: Tree Representation of a Set, Weighted Union and Path Compression-Analysis and Applications.	
GREEDY DESIGN TECHNIQUE	**(06 Hours)**
Review of Basic Greedy Control Abstraction, Activity Selection Problem & Variants, Huffman Coding, Horn	

Formulas; The Knapsack Problem, Clustering; Minimum-Cost Arborescence; Multi-phase Greedy Algorithms, Graph Algorithms; Graph problems: Graph Searching, BFS, DFS, Shortest First Search Minimum Spanning Trees, Single Source Shortest Paths, Maximum Bipartite Cover Problem, Applications, Topological Sort; Connected and Bi-connected Components; Johnson's Implementation of Prim's algorithm using Priority Queue Data Structures.	
DYNAMIC PROGRAMMING	**(08 Hours)**
The Coin Changing Problem, The Longest Common Subsequence, The 0/1 Knapsack Problem; Memoization; Dynamic Programming over Intervals, Shortest Paths and Distance Vector Protocols; Constructing Optimal Binary Search Trees; Algebraic Problems: Evaluation of Polynomials With or Without Preprocessing; Winograd's and Strassen's Matrix Multiplication Algorithms and Applications to Related Problems, FFT, Simple Lower Bound Results.	
STRING PROCESSING	**(02 Hours)**
String Searching and Pattern Matching, Knuth-Morris-Pratt Algorithm and its Analysis; Probabilistic Algorithms, Motivation.	
BACKTRACKING AND BRANCH & BOUND	**(02 Hours)**
Backtracking, General Method, 8-Queens' Problem, Sum of Subsets Problem, Graph Coloring, Hamiltonian Cycles; Branch and Bound to Solve Combinatorial Optimization Problems.	
NP Theory	**(08 hours)**
Polynomial Time Verification, NP-Completeness & the Search Problems, The Reductions, Dealing with NP-Completeness, Local Search Heuristics, Space Complexity; Selected Topics - Algorithms for String Matching, Amortized Analysis, Bloom Filters & Their Applications.	
PROBABILISTIC ALGORITHMS	**(02 Hours)**
Indicator Random Variables, Four Main Design Categories, Randomization of Deterministic Algorithms, Monte Carlo Algorithms, Las Vegas Algorithms, Numerical Probabilistic Algorithms & Various Candidate Applications.	
APPROXIMATION ALGORITHMS	**(02 Hours)**
Introduction and Motivation for Approximation Algorithms, Greedy and Combinatorial Methods; Scheduling: Multiprocessor Scheduling.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on website.)	
1	Lab assignments based on designing algorithms for trivial computational problems and doing their

	empirical timing analysis.
2	Lab assignments based on designing algorithms using divide and conquer technique and doing their empirical timing analysis.
3	Lab assignments based on designing algorithms using greedy technique and doing their empirical timing analysis.
4	Lab assignments based on designing algorithms using dynamic programming and doing their empirical timing analysis.
5	Lab assignments based on backtracking & branch bound approach to design algorithms.
6	Lab assignments based on designing Approximation algorithms to solve the hard computational problems.

BOOKS RECOMMENDED (LATEST EDITION)

1. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, “Introduction to Algorithms”, The MIT Press.
2. Donald E. Knuth, “The Art of Computer Programming, Vol. 1, Vol. 2 and Vol. 3”, Narosa/Addison Wesley, New Delhi/London.
3. Ellis Horowitz, SartajSahni, “Data Structures, Algorithms and Applications in C++”, Universities Press/Orient Longman.
4. J. Kleinberg, E. Tardos, “Algorithm Design”, Pearson Education.
5. Sara Baase, Allen V. Gelder, “Computer Algorithms”, Pearson Education.

ADDITIONAL BOOKS RECOMMENDED

1. K. Mehlhom, “Data Structures and Algorithms, Vol. 1 and Vol. 2”, Springer-Verlag, Berlin.
2. A. Borodin and I. Munro, “The Computational Complexity of Algebraic and Numeric Problems”, American Elsevier, New York.
3. Winograd, “The Arithmetic Complexity of Computation”, SIAM, New York.

Course Outcomes

At the end of the course, students will

CO1	have knowledge about the application of mathematical formula/technique to solve the computational problem.
CO2	be able to understand, identify and apply the most appropriate algorithm design technique required to solve a given problem.
CO3	be able to analyze and compare the asymptotic time and space complexities of algorithms.

CO4	be able to write rigorous correctness proofs or implementation for algorithms.
CO5	be able to design and give the solution using innovate/synthesize algorithms to solve the computational problems.

M. Tech. – I Semester – I	L	T	P	C
CSEDS605: MACHINE LEARNING (CORE-3)	**3**	**0**	**2**	**4**

Course Objective	
1	To understand the basic concepts, state-of-the art techniques of machine learning, statistical analysis and discriminant functions.
2	To apply different concepts for the machine learning problems.
3	To analyze supervised and unsupervised learning approaches as per the suitability of the problem.
4	To evaluate machine learning methods for performance and usage for different problems.
5	To design solution of problem using different machine learning approaches.

INTRODUCTION	**(04 Hours)**
Pattern Representation, Concept of Pattern Recognition, Basics of Probability, Bayes' Decision Theory, Maximum-Likelihood and Bayesian Parameter Estimation, Error Probabilities, Learning of Patterns, Modeling, Regression, Discriminant Functions, Linear Discriminant Functions, Decision Surface, Learning Theory, Fisher Discriminant Analysis.	
SUPERVISED LEARNING ALGORITHMS	**(06 Hours)**
Gradient Descent, Linear Regression, Support Vector Machines, K-Nearest Neighbor, Naïve Bayes, Bayesian Networks, Classification, Decision Trees, ML and MAP Estimates, Overfitting, Regularization, Bayes Classification, Nearest Neighbor Classification, Cross Validation and Attribute Selection, Bayesian Decision Theory, Losses and Risks, Bayesian Networks, Parametric Methods: Gaussian Parameter Estimation, Maximum Likelihood Estimation, Bias and Variance, Bayes' Estimator, Bayesian Estimation, Parametric Classification, Regression, Naive Bayes, Hidden Markov Models, Support Vector Machines, Decision Trees.	
NEURAL NETWORKS AND LEARNING ALGORITHMS	**(08 Hours)**
Artificial Neural Networks, Perceptron, Multilayer Networks, Back Propagation, Deep Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks; Linear Discrimination, Multilayer Perceptron: Multilayer Perceptron, Backpropagation Algorithm, Nonlinear Regression, Convergence, Overtraining, Dimensionality Reduction, Gradient Descent, Recurrent Networks, Cross-Validation and Resampling Methods, Bootstrapping.	
UNSUPERVISED LEARNING ALGORITHMS	**(06 Hours)**
Nonparametric Methods: Nonparametric Density Estimation, Histogram Estimator, Kernel Methods, Properties of Kernels, Kernel Estimator, K-Nearest Neighbor Estimator, Nonparametric Classification, K-	

Means Clustering, Gaussian Mixture Models, Learning with Partially Observable Data, Expectation Maximization Algorithm.	
MISCELLANEOUS TOPICS	**(08 Hours)**
DimensionalityMeasuring Error, Interval Estimation, Hypothesis Testing, Reduction, Feature Selection, Principal Component Analysis, Pattern Analysis using Eigen Decomposition, Principal Component Analysis, Parzen-windows Method, Model Selection and Theory of Generalization, In-sample and Out-of-sample Error, Vapnik-Chervonenkis (VC) Dimension, VC Inequality, VC Analysis.	
APPLICATIONS	**(10 Hours)**
Signal Processing, Image Processing, Biometric Recognition, Face and Speech Recognition, Information Retrieval, Natural Language Processing.	
Practical and mini-projects will be based on the coverage of the above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Implement classification and regression techniques.
2	Implement clustering and statistical modeling methods.
3	Implement various dimensionality reduction techniques.
4	Implement neural networks and non-parametric techniques.
5	Implement mini-project based on machine learning approaches.

BOOKS RECOMMENDED (LATEST EDITION)

1. Richard O. Duda, Peter E. Hart, David G. Stork, “Pattern Classification”, Wiley.
2. Christopher M. Bishop, “Pattern Recognition and Machine Learning”, Springer.
3. Geoff Dougherty, “Pattern recognition and classification an Introduction”, Springer.
4. Richard O. Duda and Peter E. Hart, “Pattern Classification and Scene Analysis”, John Wiley & Sons.
5. John Shae Taylor and Nello Cristianini, “Kernel Methods for Pattern Analysis” Cambridge University Press.

ADDITIONAL BOOKS RECOMMENDED

1. Ranjjan Shinghal, “Pattern Recognition Techniques and Application”, Oxford University Press.
2. Theodoridis and K. Koutroumbas, “Pattern Recognition”, Academic Press.
3. Judith L. Gersting, “Mathematical Structure for Computer Science”, W.H. Freeman and Co.

Course Outcomes **At the end of course, students will**	
CO1	have knowledge of pattern recognition, regression, classification, clustering algorithms and statistics.
CO2	be able to apply different feature extraction, classification, regression, neural network algorithms and modeling.
CO3	be able to analyze the data patterns and modeling for applying the learning algorithms and non-parametric approaches.
CO4	be able to evaluate the performance of an algorithm and comparison of different learning techniques.
CO5	be able to design solutions for real life problems like biometric recognition, natural language processing, and related applications using various tools and techniques of machine learning.

M. Tech. – I Semester – I	L	T	P	C
CSE613: FOUNDATIONS OF DATA SCIENCE (CORE-4)	**3**	**0**	**2**	**4**

Course Objective	
1	To understand the fundamentals of data analytics, distributed database, foundational skills in data science, including preparing and working with data; abstracting and modeling.
2	To go from raw data to a deeper understanding of the patterns and learn to store, manage, and analyze unstructured data structures within the data, to support making predictions and decision making.
3	To learn processing large data sets using Hadoop and make predictions using machine learning and statistical methods.
4	To learn computational thinking and skills, various text analysis and stream data analysis techniques including the Python programming language for analyzing and visualizing data.
5	To learn various topics such as statistics, crawling data, data visualization, advanced databases, complex data represented using graphs or high dimensional data and cloud computing, along with a toolkit to use with data.

INTRODUCTION	**(06 hours)**
Overview of Data Science and Big Data, Datafication: Current landscape of Perspectives, Skill Sets needed; Matrices, Matrices to Represent Relations Between Data and Linear Algebraic Operations on Matrices, Approximately Representing Matrices by Decompositions, SVD and PCA; Statistics: Descriptive Statistics: Distributions and Probability, Statistical Inference: Populations and Samples, Statistical Modeling, Fitting a Model, Hypothesis Testing, Introduction to R and Python.	
DATA PREPROCESSING	**(08 hours)**
Types of Data and Representations, Acquiring Data, Crawling, Parsing Data, Data Manipulation, Data Wrangling, Data Cleaning, Data Integration, Data Reduction, Data Transformation, Data Discretization, Distance Metrics, Evaluation of Classification, Methods: Confusion Matrix, Student's T-tests and ROC Curves, Exploratory Data Analysis, Basic Tools: Plots, Graphs and Summary Statistics of EDA, Philosophy of EDA.	
GRAPH	**(08 Hours)**
Different Types of Graphs, Trees, Basic Concepts Isomorphism and Subgraphs, Multi Graphs and Euler Circuits, Hamiltonian Graphs, Chromatic Numbers, Graph and Tree Processing Algorithms, Graph based Applications	

DATA VISUALIZATION	**(04 hours)**
Data visualization: Basic Principles and Tools, Graph Visualization, Data summaries, Link analysis, Mining of Graph, High Dimensional Clustering, Recommendation Systems.	
PARADIGMS FOR LARGE SCALE DATA PROCESSING	**(08hours)**
MapReduce, Hadoop System, Software Interfaces, e.g., Hive, Pig, Traditional Warehouses vs. MapReduce Technology, Distributed Databases, Distributed Hash Tables, Near-real-tips Query.	
TEXT ANALYSIS	**(08 hours)**
Data Flattening, Filtering, Chunking, Feature Scaling, Dimensionality Reduction, Nonlinear Futurization, Shingling of Documents, Locality-Sensitive Hashing for Documents, Distance Measures, LSH Families for Other Distance Measures, Collaborative Filtering, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream, Moments, Windows, Clustering for Streams.	
Practical will be based on the coverage of the above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Practical related to Hadoop Installation and implementations using artificial data.
2	Introduction to software tools for data analytics science.
3	Practical based on Basic Statistics and Visualization.
4	Practical related to data preprocessing and data preparation for various Data mining processes.
5	Practical related to different SQL and NOSQL databases.
6	Practical based on Classification.
7	Practical based on K-means Clustering.
8	Practical related to Big Text analysis.

BOOKS RECOMMENDED (LATEST EDITION)

1. Joel Grus, “Data science from scratch”, O'Reilly Media.
2. Avrim Blum, John Hopcroft, and Ravindran Kannan, “Foundations of Data Science”, Cambridge University Press.
3. Anand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, Cambridge University Press.
4. Peter Bruce, Andrew Bruce, ”Practical Statistics for Data Scientists: 50”, O’Reilly publishing house.
5. Douglas C. Montgomery and George C. Runger, “Applied statistics and probability for engineers”, John Wiley & Sons.

ADDITIONAL BOOKS RECOMMENDED
1. Jiawei Han, Micheline Kamber and Jian Pei, “Data Mining: Concepts and Techniques”, Morgan Kaufmann. 2. Mohammed J. Zaki and Wagner Miera Jr, “Data Mining and Analysis: Fundamental Concepts and Algorithms”, Cambridge University Press. 3. Matt Harrison, “Learning the Pandas Library: Python Tools for Data Munging, Analysis, and Visualization, O'Reilly. 4. Tom White, ”Hadoop: The Definitive Guide”, O’Reilly Media.

Course Outcomes **At the end of the course, students will**	
CO1	be able to understand the principles and purposes of data science, and articulate the different dimensions of the area.
CO2	be able to apply various data pre-processing and manipulation techniques including various distributed analysis paradigms using Hadoop and other tools.
CO3	be able to apply basic data mining machine learning techniques to build a classifier or regression model, and predict values for new examples.
CO4	be able interpret various large datasets by applying Data Mining techniques like clustering, filtering, factorization.
CO5	be able to implement and perform advanced statistical analysis to solve complex and large dataset problems for real life applications.

M. Tech. – I Semester – I	L	T	P	C
CSE607: RESEARCH METHODOLOGY IN CSE	**4**	**0**	**0**	4

Course Objective	
1	To understand the basic terminology of research, its methodology and learn different methodologies of pursuing the research in terms of organization, presentation and evaluation.
2	To apply the concept in writing the technical content.
3	To analyze the existing method using different parameters in different scenarios.
4	To evaluate the proposed work and compare with existing approach systematically using the appropriate methodology, through simulation depending upon the research field.
5	To design algorithms using concepts learned and write report and papers technically and grammatically correct.

INTRODUCTION	**(06 Hours)**
Research: Definition, Characteristics, Motivation and Objectives, Research Methods vs Methodology, Types of Research – Descriptive vs Analytical, Applied vs Fundamental, Quantitative vs Qualitative, Conceptual vs Empirical.	
METHODOLOGY	**(05 Hours)**
Research Process, Formulating the Research Problem, Defining the Research Problem, Research Questions, Research Methods vs. Research Methodology.	
LITERATURE REVIEW	**(05 Hours)**
Review Concepts and Theories, Identifying and Analyzing the Limitations of Different Approaches.	
FORMULATION AND DESIGN	**(06 Hours)**
Concept and Importance in Research, Features of a Good Research Design, Exploratory Research Design, Concept, Types and Uses, Descriptive Research Designs, Concept, Types and Uses, Experimental Design: Concept of Independent & Dependent Variables.	
DATA MODELING AND SIMULATIONS	**(08 Hours)**
Mathematical Modeling, Experimental Skills, Simulation Skills, Data Analysis and Interpretation.	
TECHNICAL WRITING AND TECHNICAL PRESENTATIONS	**(08 Hours)**
CREATIVITY AND ETHICS IN RESEARCH, INTELLECTUAL PROPERTY RIGHTS	**(04 Hours)**
TOOLS AND TECHNIQUES FOR RESEARCH	**(06 Hours)**
Methods to Search Required Information Effectively, Reference Management Software, Software	

for Paper Formatting, Software for Detection of Plagiarism.	
DISCUSSION AND DEMONSTRATION OF BEST PRACTICES	**(08 Hours)**
(Total Contact Time: 56 Hours)	

BOOKS RECOMMENDED (LATEST EDITION)

1. John W. Creswell, “Research Design: Qualitative, Quantitative, and Mixed Methods Approaches”, SAGE Publications Ltd.
2. C.R. Kothari, ”Research Methodology: Methods and Techniques”, New Age International Publishers.
3. David Silverman, ”Qualitative Research”, SAGE Publications Ltd.
4. Norman K. Denzin and Yvonna Sessions Lincoln, ”Handbook of Qualitative Research”, SAGE Publications Ltd.
5. Michael Quinn Patton, ”Qualitative Research and Evaluation Methods”, SAGE Publications Ltd.

Course Outcomes **At the end of the course, students will**	
CO1	have an understanding of the different research methodology in different areas.
CO2	be able to apply the concepts in writing, presentation, and simulating different experiments.
CO3	be able to analyze the proposed work with existing approaches in the literature and interpret the research design through project development and case study analysis using appropriate tools.
CO4	be able to execute the technical presentation, organization in writing the report and papers.
CO5	be able to design the algorithms and proof learned and communicate effectively through proper organization and presentation.

M. Tech. – I Semester – I	L	T	P	C
CSEDS611: INFORMATION RETRIEVAL (CORE ELECTIVE-1)	**3**	**0**	**2**	**4**

Course Objective	
1	To understand the basic building blocks of information retrieval systems.
2	To introduce a variety of indexing techniques, retrieval models and ranking algorithms for information retrieval.
3	To provide comprehensive details of evaluation methods used for information retrieval systems.
4	To apply classification and clustering approaches for information retrieval.
5	To introduce the basic concepts of web information retrieval.

INTRODUCTION	**(04 Hours)**
Information Retrieval Problem, Unstructured and Semi-structured Data, Inverted Index, Processing Boolean Queries, Posting Lists and Dictionaries.	
INDEX CONSTRUCTION AND COMPRESSION	**(10 Hours)**
Sort-Based Index Construction, Hardware Basics, Blocked Sort-Based Indexing, Single-Pass In-Memory Indexing, Distributed Indexing, Dynamic Indexing, Other Types of Indexes such as Positional Indexes and N-Gram Indexes, Statistical Properties of Terms: Heaps’ Law and Zipf’s Law, Dictionary Compression, Postings Compression.	
RETRIEVAL MODELS AND SCORING	**(10 Hours)**
Boolean, Vector Space, Probabilistic and Semantic Modeling, Vector Space Scoring, TF IDF Weighting, Inverse Document Frequency, The Cosine Measure, Efficient Scoring and Ranking in Search Systems, Relevance Feedback and Query Expansion.	
EVALUATION IN INFORMATION RETRIEVAL SYSTEM	**(06 Hours)**
Standard Test Collections, User Happiness, Precision, Recall, F-Measure, Unranked Retrieval Sets and Ranked Retrieval Results Evaluation, Assessing Relevance, System Quality and User Utility: A Broader Perspective.	
TEXT CLASSIFICATION AND CLUSTERING	**(08 Hours)**
Introduction to Text Classification, Naive Bayes Text Classification, Vector Space Classification (Using Hyper planes, Centroids and K Nearest Neighbors), Support Vector Machine Classifiers, Clustering vs Classification, Partitioning Methods, K-Means Clustering, Hierarchical Clustering.	
OTHER TOPICS IN INFORMATION RETRIEVAL	**(04 Hours)**
Web Crawling, Search Engines, Ranking, Link Analysis, Page Rank, XML Retrieval, Semantic Web.	

Practical and mini-projects will be based on the coverage of the above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practicals (Problem statements will be changed every year and will be notified on the website.)	
1	Implementation of sort-based and single-pass in-memory indexing.
2	Implementation of distributed and dynamic indexing.
3	Implementation of n-gram indexes.
4	Programs to demonstrate boolean retrieval and vector space models.
5	Program to find the similarity between documents.
6	Implementation of naive bayes text classification.
7	Implementation of vector space classification algorithms such as k nearest neighbor.
8	Programs to implement k-means clustering and hierarchical clustering.
9	Implementation of page rank algorithm.
10	Mini project.

BOOKS RECOMMENDED (LATEST EDITION)

1. Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, “Introduction to Information Retrieval”, Cambridge University Press.
2. Stefan Buttcher, Charlie Clarke, Gordon Cormack, “Information Retrieval: Implementing and Evaluating Search Engines”, The MIT Press.
3. Bruce Croft, Donald Metzler, Trevor Strohman, “Search Engines: Information Retrieval in Practice”, Pearson Education.
4. Baeza-Yates Ricardo, Berthier Ribeiro-Neto, “Modern Information Retrieval”, Addison-Wesley.
5. Christopher M. Bishop, “Pattern Recognition and Machine Learning”, Springer.

Course Outcomes **At the end of the course, students will**	
CO1	be able to understand different information retrieval models and indexing techniques.
CO2	understand different text compression algorithms and their role in efficient building and storage of inverted indexes.
CO3	know about different evaluation methods used for information retrieval systems.
CO4	be able to understand the application of various classification and clustering techniques for information retrieval systems.

CO5	be able to understand the working of a search engine and the page ranking algorithm.
CO6	know about the basics of XML retrieval and web search.

M. Tech. I Semester – I	L	T	P	C
CSEDS613: ADVANCED DATABASE MANAGEMENT SYSTEMS (CORE ELECTIVE-1)	**3**	**0**	**2**	**4**

Course Objective	
1	Enhanced the knowledge in the areas of database management that go beyond traditional (relational) database management systems.
2	Comprehend the query processing efficient information management for Distributed , Parallel and Object Oriented DBMS.
3	To understand and implement of different data and their database management systems.
4	To enhance the knowledge about variety of data storage and management.
5	To understand storage and management issues of the unstructured data.

DISTRIBUTED DATABASE CONCEPTS	**(6 Hours)**
Overview of client - server architecture and its relationship to distributed databases, Concurrency control Heterogeneity issues, Persistent Programming Languages, Object Identity and its implementation, Clustering, Indexing, Client Server Object Bases, Cache Coherence.	
PARALLEL DATABASES	**(6 Hours)**
Parallel Architectures, performance measures, shared nothing/shared disk/shared memory based architectures, Data partitioning, Intra-operator parallelism, Pipelining, Scheduling, Load balancing	
QUERY PROCESSING	**(6 Hours)**
Index based, cost estimation, Query optimization: algorithms, Online query processing and optimization, XML, DTD, XPath, XML indexing, Adaptive query processing.	
ADVANCED TRANSACTION MODELS	**(6 Hours)**
Savepoints, Sagas, Nested Transactions, Multi Level Transactions. Recovery: Multilevel recovery, Shared disk systems, Distributed systems 2PC, 3PC, replication and hot spares, Data storage, security and privacy Multidimensional K- Anonymity, Data stream management.	
MODELS OF SPATIAL DATA	**(5 Hours)**
Conceptual Data Models for spatial databases (e.g. pictogram enhanced ERDs), Logical data models for spatial databases: raster model (map algebra), vector model, Spatial query languages, Need for spatial operators and relations, SQL3 and ADT. Spatial operators, OGIS queries	
WEB ENABLED APPLICATIONS	**(5 Hours)**

Review of 3-tier architecture - Typical Middle-ware products and their usage. Architectural support for 3 -tier applications: technologies like RPC, CORBA, COM. Web Application server - WAS architecture Concept of Data Cartridges - JAVA/HTML components. WAS	
OBJECT ORIENTED DATABASES	**(4 Hours)**
Notion of abstract data type, object oriented systems, object oriented db design. Expert databases: use of rules of deduction in data bases, recursive rules.	
ADVANCED TOPICS	**(4 Hours)**
No SQL Databases, Unstructured Databases, Couchbase, MangoDB, Cassendra, Redis, Memcached.	
Practical will be based on the coverage of the above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Write queries and analyze the query performances.
2	Implementation of problem having spatial data.
3	Implement the web based application with database connectivity.
4	Implementation of problem using object oriented concept.
5	Analyse the performance of problem using row oriented database vs no SQL databases.
6	Optimization of Distributed Database Queries: as a Mini Project.

BOOKS RECOMMENDED (LATEST EDITION)

1. R. Elmasriand S. Navathe, Fundamentals of Database Systems, Benjamin- Cummings.
2. AviSilberschatz, Hank Korth, and S. Sudarshan, Database System Concepts, McGraw Hill.
3. S. Shekhar and S. Chawla, Title Spatial Databases: A Tour, Prentice Hall.
4. Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom, Database Systems, Pearson.
5. Mattison, Rob Mattison, "Web Data Warehousing and Knowledge Management", MGH.
6. W. Kim, "Introduction to Object Oriented Databases", MIT Press.

Course Outcomes **At end of the course Student will be able to**	
CO1	Understand advanced database techniques for storing a variety of data with various database models.
CO2	To apply various database techniques/functions with Object Oriented approach to design database for real life scenarios.

CO3	Analyse the problem to design database with appropriate database model.
CO4	Evaluate methods of storing, managing and interrogating complex data.
CO5	Implement web application API’s, distributed databases with the integration of various programming languages.

M. Tech. – I Semester – I	L	T	P	C
CSEDS615: EMBEDDED SYSTEMS DESIGN (CORE ELECTIVE-1)	**3**	**0**	**2**	**4**

Course Objective	
1	To learn about hardware and software design requirements of embedded systems, theprocesses, methodologies, fundamental problems, and best practices associated with the development of applications in the context of high-performance embedded computing systems.
2	To study several different styles of processors used in embedded systems, the use ofinterrupts and inter-process communication, techniques for tuning the performance of a processor, and to optimize embedded CPUs.
3	To understand memory system optimizations and the back end of the compilation process to determine the quality of code.
4	To study the importance of embedded multiprocessors, their architectures, design techniques, methodologies, algorithms, IoT, and its applications.
5	To learn various embedded software development tools and provide in-depth knowledge of scheduling algorithms and middleware architectures for multiprocessors and hardware/software co-design and co-synthesis algorithms.

INTRODUCTION: EMBEDDED HARDWARE	**(04 Hours)**
Introduction to embedded systems Hardware needs; typical and advanced, timing diagrams, memories (RAM, ROM, and EPROM) Tristate devices, Buses, DMA, UART and PLD's Built-ins on the microprocessor, Example applications, Design methodologies, Embedded Systems Design flows, Models of computation, Parallelism and computation, Reliable system design, CE architecture.	
INTERRUPTS	**(04 Hours)**
Interrupts basics ISR; Context saving, shared data problem. Atomic and critical section, Interrupt latency.	
SOFTWARE AND OS	**(04 Hours)**
Survey of software architectures, Round Robin, Function queue scheduling architecture, Use of real time operating system, RTOS, Tasks, Scheduler, Shared data reentrancy, priority inversion, mutex binary semaphore and counting semaphore, Parallel execution mechanisms, Superscalar, SMID and Vector processors, Variable performance CPU architectures, CPU Simulation, Automated CPU Design.	
INTER-PROCESS COMMUNICATION	**(05 Hours)**
Inter task communication, message queue, mailboxes and pipes, timer functions, events Interrupt routines	

in an RTOS environment.	
EMBEDDED COMPUTING	**(07 Hours)**
Embedded design process, System description formalisms, Instruction sets- CISC and RISC, DSP processors, Embedded computing platform- CPU bus, Memory devices, I/O devices, interfacing, designing with microprocessors, debugging techniques, Hardware accelerators- CPUs and accelerators, Accelerator system design, Embedded system software design using an RTOS Hard real-time and soft real-time system principles, Task division, need of interrupt routines, shared data.	
INTERNET OF THINGS	**(04 Hours)**
Introduction, IoT work flow, IoT Protocols: HTTP, CoAP, MQTT, 6 LoWPAN, building IoT applications.	
TOOLS	**(06 Hours)**
Embedded Software development tools. Host and target systems, cross compilers, linkers, locators for embedded systems. Getting embedded software in to the target system, Debugging techniques like JTAGS, Testing on host machine, Instruction set emulators, logic analyzers In-circuit emulators and monitors.	
NETWORK	**(04 Hours)**
Distributed embedded architectures, Networks for embedded systems, Network-based design, and Internet enabled systems.	
SYSTEM DESIGN TECHNIQUES	**(04 Hours)**
Design methodologies, Requirements analysis, System analysis and architecture design, Quality assurance.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on website.)	
1	Implement experiment based on programming of Embedded boards.
2	Implement experiment based on Embedded OS.
3	Implement RTOS and job scheduler with Embedded systems.
4	Implement Embedded computing algorithm and evaluate the performance using different tools.
5	Implement mini projects based on Embedded systems for real applications.

BOOKS RECOMMENDED (LATEST EDITION)

1. Mohamed Ali Mazidi, Janice GillispieMazidi, RolinMcKinlay, “The 8051 Microcontroller and Embedded Systems: Using Assembly and C”, Pearson Education.
2. Raj Kamal, “Embedded Systems-Architecture, Programming and Design”, TMH.
3. Jonathan W. Valvano, “Embedded Microcomputer Systems-Real Time Interfacing”, Thomson Learning.
4. David A. Simon, “An Embedded Software Primer”, Pearson Education.
5. Louis L. Odette, “Intelligent Embedded Systems”, Addison-Wesley.

ADDITIONAL BOOKS RECOMMENDED

1. Wayne Wolf, “High-Performance Embedded Computing: Architectures, Applications, and Methodologies“, Morgan Kaufmann.
2. Larry L Peterson, “Computer Networks: A Systems Approach”, Morgan Kaufmann.
3. Frank Vahid and Tony Givargis, “Embedded System Design: A Unified Hardware/Software Introduction”, John Wiley.
4. Marilyn Wolf, “Computers as Components- Principles of Embedded Computing System Design”, Morgan Kaufmann.
5. Denial D. Gajski , Frank Vahid, “Specification and design Embedded systems”, Prentice Hall; Facsimile edition.

Course Outcomes **At the end of the course, students will**	
CO1	be able to understand hardware-software requirements, interrupts and inter process communication of embedded systems.
CO2	be able to apply techniques for simulating processors, for tuning the performance of a processor and to optimize embedded CPUs, such as code compression and bus encoding. They will be able to use middleware architectures for dynamic resource allocation in multiprocessors
CO3	be able to analyze the embedded systems’ specifications and develop software programs.
CO4	be able to evaluate related software architectures and tools for embedded Systems and evaluate the quality of code using the back end of the compilation process and be able to characterize embedded applications and target architectures using different models.
CO5	be able to design and develop real time embedded systems using the concepts of RTOS.

M. Tech. – I Semester – II	L	T	P	C
CSEDS617: COMPUTER VISION AND IMAGE PROCESSING (CORE ELECTIVE- 1)	**3**	**0**	**2**	**4**

Course Objective	
1	To learn how to use Data science for Image Processing and Vision tasks.
2	To understand the capability of machines to analyze visual information and to make appropriate decisions.
3	To learn various methods and algorithms for Image Processing and Computer Vision.
4	To learn applications with a combination of Image/Vision, Machine Learning and Artificial Intelligence.
5	To understand various applications of Image and Computer Vision.

INTRODUCTION	**(06 Hours)**
Introduction, Motivation, Introduction to Image Formation, Capture and Representation, Linear Filtering, Correlation, Convolution, Image Recognition Applications.	
IMAGE REPRESENTATION AND TECHNIQUES	**(08 Hours)**
Introduction, Image Digitization, Discrete Fourier Transform, Image Pre-Processing in Spatial and Frequency Domain, Conventional Image Processing Techniques, Local Pre-Processing and Global Pre-Processing.	
IMAGE SEGMENTATION	**(06 Hours)**
Segmentation Techniques, Object Segmentation, Identification of Objects, Object Detection and Semantics Segmentation with CNNs.	
ATTENTION MODELS	**(10 Hours)**
Introduction to Attention Models in vision, Vision and Language, Image Captioning, Visual QA, Visual Dialog, Spatial Transformers, Transformer Networks.	
APPLICATIONS OF VISION	**(12 Hours)**
Communication through Vision, Detection and Recognition, Vision Understanding, Scene Understanding, Inference and Decision Making, Video Image Characteristics, Classification of Images, Image Generation.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Installation and working on OpenCV, Scilab etc.
2	Learning Hadoop systems for implementing Data science applications of image processing and vision.

3	Development of applications using pytorch library.
4	Use of Machine learning and deep learning techniques for solving image and/or vision related problems.
5	Comparative evaluation of deep learning models for image and vision.

BOOKS RECOMMENDED (LATEST EDITION)

1. Richard Szeliski, “Computer Vision: Algorithms and Applications”, Springer.
2. Simon Prince, “Computer Vision: Models, Learning and Inference”, Cambridge University Press.
3. Rafael C. Gonzalez, Richard E. Woods, “Digital Image Processing”, Pearson Education.
4. Matthew Turk, Gang Hua, “Vision-based Interaction”, Morgan Claypool.
5. Ian Good fellow, YoshuoBengio, Aaron Courville, “Deep Learning (Adaptive Computation and Machine Learning series)”, The MIT Press.

Course Outcomes

At the end of the course, students will

CO1	have knowledge about various methods and algorithms for image processing and computer vision.
CO2	be able to apply algorithms and methods for large datasets for image and vision.
CO3	be able to apply image and vision algorithms in SciLab/OpenCV for applications.
CO4	be able to apply image and vision based solutions for specific real-world applications.
CO5	be able to analyze data science techniques for image and video processing.

M. Tech. – I Semester – I	L	T	P	C
CSEDS619: SPEECH AND AUDIO PROCESSING (CORE ELECTIVE-1)	**3**	**0**	**2**	**4**

Course Objective	
1	To learn the basics of digital signal processing, analytical methods and it's different applications
2	To understand fundamentals of speech
3	To learn different speech models and speech processing
4	To learn the design of different filters in spatial and frequency domain for speech processing
5	To develop skills for analyzing and synthesizing algorithms and systems for speech recognition, identification, classification for different applications.

BASICS OF DIGITAL SIGNAL	**(06 Hours)**
Analog vs. Digital Signal, Continuous vs. Discrete Signal, Issues with Analog signal processing, Digital signal transmission, Overview of different applications, Fundamentals of z-transform, Fourier transform, Overview of Digital filters: FIR and IIR, Sampling theorem, Decimation and Interpolation.	
FUNDAMENTALS OF SPEECH	**(04 Hours)**
Speech signal, Digital representation of speech, Speech production and perception, Acoustic modeling, Acoustic tubes and features, Acoustic phonetics, Sound propagation, Phase vocoder, Channel vocoder, Vocal tract functioning, Vocal tract transfer function, Time domain models, Frequency domain representation, Concepts of Subband.	
TIME DOMAIN ANALYSIS	**(08 Hours)**
Short time energy and average magnitude, Short time average zero-crossing rate, Pitch period estimation, Speech and silence discrimination, Short time autocorrelation function, Median smoothing, Quantization, Companding, Adaptive Quantization, Delta modulation, Differential PCM.	
FREQUENCY DOMAIN ANALYSIS	**(08 Hours)**
Short time Fourier representation, Short time analysis, Spectrographic, Spectrum analysis, Complex Cepstrum, Pitch Detection, Formant estimation, Linear predictive analysis, LPC equation, solutions, Frequency domain interpretation of Linear Predictive analysis, Relations between various speech parameters, Applications of LPC parameters, IIR and FIR filters design.	
SPEECH MODELING AND PROCESSING	**(16 Hours)**
Vocabulary, Language Modeling, Hidden Markov Models, Pattern Classification and Recognition, Speech Compression, Speech synthesis, Speech recognition, Speaker identification, Emotion analysis, Language identification, Speech Conversion, Speech processing using Neural Networks, Deep Learning.	

Practical and mini-projects will be based on the coverage of the above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on website.)	
1	Implementation of basic signal transforms like Fourier, Wavelet and others.
2	Implementation of preliminary feature extractions from speech signals.
3	Implementation of time domain analysis techniques and design of different filters.
4	Implementation of frequency domain analysis techniques and design of different filters.
5	Implementation of advanced techniques of modelling for speech processing.
6	Implementation of application based mini project.

BOOKS RECOMMENDED (LATEST EDITION)

1. Lawrence R. Rabiner and Ronald W. Schafer, “Theory and Applications of Digital Signal Processing”, Pearson.
2. Lawrence R. Rabiner and Ronald W. Schafer, “Digital Processing of Speech Signals”, Pearson.
3. Lawrence Rabiner, Biing-Hwang Juang, B. Yegnanarayana, “Fundamentals of Speech Recognition”, Pearson.
4. Douglas O’Shaughnessy, “Speech Communications Human and Machines”, Institute of Electrical and Electronics Engineers.
5. Ben Gold and Nelson Morgan, “Speech and Audio Signal Processing”, Wiley.

ADDITIONAL BOOKS RECOMMENDED

1. M. R. Schroeder, “Computer Speech: Recognition, Compression, Synthesis”, Springer Series in Information Science.

Course Outcomes **At the end of the course, students will**	
CO1	be able to understand the process of converting the continuous-time signal into digital signal, process it and convert back to continuous-time signal
CO2	be able to apply the different digital filters to design speech processing applications
CO3	be able to analyse the speech in time domain and frequency domain and also able to analyse tools like Fourier transform and z-transform to find a system's frequency response or system's impulse response
CO4	be able to evaluating the performance of a speech processing based systems like speech recognition,

	speech identification and many more
CO5	be able to design robust and efficient the speech models and speech processing systems

M. Tech. – I Semester – I	L	T	P	C
CSEDS621: High-Performance Computing (CORE ELECTIVE-1)	**3**	**0**	**2**	**4**

Course Objective	
1	To understand fundamentals concepts related to High-Performance Computing and state-of-the-art in Parallel Programming environment
2	To study the architectures of several types of high-performance computers and the implications on the performance of algorithms of these architectures
3	To provide an in-depth analysis of design issues in parallel computing
4	To learn the programming constructs required for parallel programming
5	To learn how to achieve parallelism in CUDA architectures

Parallel Processing Concepts	**(10 Hours)**
Levels of parallelism (instruction, transaction, task, thread, memory, function), Models (SIMD, MIMD, SIMT, SPMD, Dataflow Models, and Demand-driven Computation etc.), Architectures: N-wide superscalar architectures, multi-core, multi-threaded, performance file systems, GPU systems, performance clusters.	
Design Issues and challenges in Parallel Computing	**(10 Hours)**
Synchronization, Scheduling, Job Allocation, Job Partitioning, Dependency Analysis, Mapping Parallel Algorithms onto Parallel Architectures, Performance Analysis of Parallel Algorithms, Bandwidth Limitations, Latency Limitations, Latency Hiding/Tolerating Techniques and their limitations, Power-Aware Computing and Communication, Power-aware Processing Techniques, Power-aware Memory Design, Power-aware Interconnect Design, Software Power Management.	
Parallel Programming with OpenMP and mpi	**(10 Hours)**
Programming languages and programming-language extensions for HPC, Inter-process communication, Synchronization, Mutual exclusion, Basics of parallel architecture, Parallel programming with OpenMP and (Posix) threads, Message passing with MPI, Thread Management, Workload Manager, Job Schedulers.	
Parallel Programming with CUDA	**(08 Hours)**
Processor Architecture, Interconnect, Communication, Memory Organization, and Programming Models in high-performance computing architectures: (Examples: IBM CELL BE, Nvidia Tesla GPU, Intel Larrabee Micro architecture and Intel Nehalem micro architecture), Memory hierarchy and transaction-specific memory design, Thread Organization, OpenCL.	
Advanced Topics	**(04 Hours)**

Peta scale Computing, Optics in Parallel Computing, Quantum Computers.	
Practical and mini-projects will be based on the coverage of the above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Implement parallel programming preliminary examples.
2	Implement algorithms using OpenMP and MPI.
3	Implement experiments using CUDA.
4	Implement and evaluate performance HPC algorithms for load distribution, thread management and job scheduling.
5	Implementation of mini-projects in different areas.

BOOKS RECOMMENDED (LATEST EDITION)

1. John L. Hennessy and David A. Patterson "Computer Architecture -- A Quantitative Approach", 4th Ed., Morgan Kaufmann Publishers.
2. Barbara Chapman, Gabriele Jost and Ruud van der Pas, "Using OpenMP: portable shared memory parallel programming", The MIT Press.
3. Marc Snir, Jack Dongarra, Janusz S. Kowalik, Steven Huss-Lederman, Steve W. Otto, David W. Walker, "MPI: The Complete Reference", Volume2, The MIT Press.
4. Pacheco S. Peter, "Parallel Programming with MPI", Morgan Kaufman Publishers.
5. Shane Cook, CUDA Programming: A Developer's Guide to Parallel Computing with GPUs, Morgan Kaufmann publishers.

Course Outcomes **At the end of the course, students will**	
CO1	be able to learn concepts, issues and limitations related to parallel computing.
CO2	be able to understand and explain different parallel models of computation, parallel architectures, interconnections and various memory organizations in modern high-performance architectures.
CO3	be able to map algorithms onto parallel architectures for parallelism.
CO4	be able to analyze and evaluate the performance of different architectures and parallel algorithms.
CO5	be able to design and implement parallel programs for shared-memory architectures and distributed-memory architectures using modern tools like OpenMP and MPI, respectively.

M. Tech. – I Semester – II	L	T	P	C
CSEDS602: ADVANCED STATISTICAL TECHNIQUES (CORE-5)	**3**	**1**	**0**	**4**

Course Objective	
1	To regain/replenish understanding core principles of statistical hypothesis testing (NHST) and techniques.
2	To use statistical principles to describe and comprehend data, including how it represents the real world and how it does not.
3	To test hypotheses or make predictions, use statistical approaches to data.
4	To create and implement analyses that result in new knowledge, decisions, and actions.
5	To communicate quantitative analysis results, conclusions and learn from and criticize the statistical analyses of others.

INTRODUCTION	**(04 Hours)**
Overview of Statistical Learning, Applications: Wage Data, Stock Market Data, Gene Expression Data, History, Statistical Learning Tools, Multivariate Approaches, Inference and Interpreting the Results of Analysis.	
STATISTICAL LEARNING	**(06 Hours)**
Statistical Learning Methods, Assessing Model Accuracy, Comparing Several Means: Analysis of Variance (ANOVA), Analysis of Covariance, Introduction to R, One-way ANOVA.	
LINEAR REGRESSION AND CLASSIFICATION	**(06 Hours)**
Simple Linear Regression, Multiple Linear Regression, Other Considerations in the Regression Model, The Marketing Plan, Comparison of Linear Regression with K-Nearest Neighbours, Logistic regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Path analysis.	
RESAMPLING METHODS	**(04 Hours)**
Bootstrapping, Cross validation, Subset Selection, Best-Subset Selection, Forward Stepwise Selection, Backward Stepwise Selection, Hybrid Methods, Dimension Reduction Methods.	
LINEAR MODEL SELECTION, REGULARIZATION AND MOVING BEYOND LINEARITY	**(08 Hours)**
PCR and PLS Regression, Polynomial Regression, Step Functions, Basis Functions, Regression Splines, Generalized Additive Models, Nonlinear Models, Factor Analysis, Multidimensional Scaling, Non-parametric techniques, Shrinkage, Ridge regression.	
TREE BASED METHODS, SUPPORT VECTOR MACHINES AND UNSUPERVISED	**(04 Hours)**
Basics of Decision Trees, Bagging, Random Forests, Boosting, Maximal Margin Classifier, Support Vector	

Classifiers, Unsupervised Learning, Principal Components Analysis, Clustering Methods.	
ADVANCED TOPICS	**(10 Hours)**
Collaborative Filtering, Pattern Matching, GeostatisticalAnaysis, Statistics in Medicine, Environmental Statistics and Causality Analysis, Efficient Statistical Sample Design	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Introduction to R, One-way ANOVA.
2	Logistic Regression, LDA, QDA, and KNN.
3	Cross-Validation and the Bootstrap
4	Non-linear Modeling.
5	Decision Trees, Lab: Support Vector Machines, Lab: PCA, Clustering, NCI60 Data Example.
6	Hands on with deep neural models.

BOOKS RECOMMENDED (LATEST EDITION)

1. Gareth James, Daniela Witten , Trevor Hastie, Robert Tibshirani, "An introduction to statistical learning" Springer.
2. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, "The elements of statistical learning", Springer.
3. Hadley Wickham and Garrett Grolemund, "R for data science", Shroff/O'Reilly.
4. Piegorsch W. Walter, "Statistical Data Analytics", John Wiley and Sons Ltd.
5. Richard Golden, "Statistical Machine Learning A Unified Framework", Taylor and Francis.

Course Outcomes **At the end of the course, students will**	
CO1	become familiar with several statistical analysis techniques.
CO2	be able to understand and analyze data in applied settings, he /she must be able to assess the appropriateness of statistical analyses, outcomes, and inferences.
CO3	be able to choose the appropriate analytical methodology for fresh research and evaluate the results accurately.
CO4	be able to learn about canonical examples of linear models to relate them to techniques and applications.
CO5	be able to conduct statistical analyses using SPDS.

M. Tech. – I Semester – II	L	T	P	C
CSEDS604: SCALABLE SYSTEMS FOR DATA SCIENCE (CORE-6)	**3**	**0**	**2**	**4**

Course Objective	
1	To understand the basic concepts and technologies of scalable distributed systems.
2	To apply the different scalable distributed system designs for solving real application problems.
3	To analyze how distributed program models such as MapReduce, TensorFlow, Vertex-centric and streaming data flows are designed to analyze large datasets.
4	To execute and compare the different popular Big Data and ML platforms like HDFS, Spark, MLLib, TensorFlow, Cassandra, Flink, etc. and understand how they are architected.
5	To develop distributed algorithms and scalable analytics applications using various design patterns.

INTRODUCTION	**(04 Hours)**
Revision of Data Structures, Arrays, Queues, Trees, Hash Maps, Graphs; Sorting Algorithms, Searching Techniques, Traversal Methods, Data Mining Basics, Statistical Limits on Data Mining.	
MEMORY-EFFICIENT DATA STRUCTURES AND APPROXIMATION	**(04 Hours)**
Memory-Efficient Data Structures, Hash Functions, Universal / Perfect Hash Families, Bloom Filters, Sketches for Distinct Count, Misra-Gries Sketch, Count Sketch, Count-Min Sketch, Approximate Near Neighbors Search, KD-Trees, LSH Families, MinHash for Jaccard, SimHash for L2, Multi-Probe, B-Bit Hashing, Data Dependent Variants, Randomized Numerical Linear Algebra, Random Projection.	
MACHINE LEARNING HARDWARE SYSTEMS	**(04 Hours)**
Machine Learning Hardware Systems, Issues, Heterogeneous Hardware Accelerators' Architecture and Accelerated Computing: Tensor Processing Units, Graphics Processing Unit.	
VIRTUAL MACHINES AND VIRTUALIZATION OF CLUSTERS AND DATA CENTRES	**(06 Hours)**
Levels of Virtualization Implementation, Design Requirements and Providers, Virtualization Support: at the OS Level and Middleware, Virtualization Tools, Hypervisor and Xen Architecture, Binary Translation with Full Virtualization, Para-Virtualization with Compiler Support, Hardware Support for Virtualization, CPU Virtualization, Memory Virtualization, I/O Virtualization, Virtualization in Multi-Core Processors, Virtual Clusters and Resource Management, Physical versus Virtual Clusters, Live VM Migration Steps and Performance Effects, Migration of Memory, Files, and Network Resources, Dynamic Deployment of Virtual Clusters.	

MAPREDUCE AND THE NEW SOFTWARE STACK	**(04 Hours)**
Distributed File Systems, MapReduce, Algorithms using MapReduce, Extensions to MapReduce, Communication Cost Model, Complexity Theory for MapReduce.	
ANALYZING BIG DATA	**(08 Hours)**
Challenges of Data Science, Introduction of Apache Spark, Data Analysis with Scala and Spark, Spark Programming Model, Record Linkage, Getting Started: The Spark Shell and Spark Context, Bringing Data from the Cluster to the Client, Shipping Code from the Client to the Cluster, Structuring Data with Tuples and Case Classes, Aggregations, Creating Histograms, Summary Statistics for Continuous Variables, Creating Reusable Code for Computing Summary Statistics, Simple Variable Selection and Scoring.	
DISTRIBUTED MACHINE LEARNING AND OPTIMIZATION	**(04 hours)**
Spark MLLib for Machine Learning: ML Algorithms, Featurization, Pipelines, Persistence, Utilities. TensorFlow for Deep Learning: Parameter Server, Federated, Alternating Direction Method of Multipliers and Applications, Clustering.	
NOSQL DATABASES AND LINKED DATA ANALYSIS	**(04 Hours)**
Consistency Models and CAP Theorem/BASE, Amazon Dynamo/Cassandra Distributed Key-Value Store, Google Big Table/HBase and SparkSQL for SQL-like Querying, Mining Social-Network Graphs, Social Networks as Graphs, Partitioning of Graphs, Finding Overlapping Communities, Simrank, Neighborhood Properties of Graphs, NOSQL Database.	
MANAGED SERVICES	**(04 Hours)**
Introduction to Cloud Computing, Cloud Strategy, Cloud Native Development, Container Adoptions, Application Modernization, Distributed App Coordination, Event Routing, Messaging, Service Discovery, Service Mesh, Workflow Orchestration,AWS, Azure.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Installation and setup of different tools mentioned in the classroom session like spark, Hadoop, HDFS, MLLib, TensorFlow, Cassandra, Flink.
2	Federated learning using edge computing and cloud computing resources, Distributed edge.
3	Experimenting with cloud storage and querying systems, Scalable querying over knowledge graphs, Scalable training and differencing over graph neural networks.
4	Experiment using Scalable pattern mining and analysis over Twitter streams.

5	Experiment using NoSQL database and application development using AWS, Azure.

BOOKS RECOMMENDED (LATEST EDITION)

1. Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, “Mining of Massive Datasets”, Cambridge University Press.
2. S. Muthukrishnan, “Data streams: Algorithms and Applications (Foundations and Trends® in Theoretical Computer Science), now Publishers Inc, USA.
3. Michael W. Mahoney, “Randomized algorithms for matrices and data: 9 (Foundations and Trends® in Machine Learning), now Publishers, USA.
4. Jimmy Lin, Chris Dyer, “Data-Intensive Text Processing with Map Reduce”, Morgan & Claypool Publishers.
5. Sandy Ryza, Uri Laserson, Josh Wills, Sean Owen, “Advanced Analytics with Spark”, O'Reilly Media Publisher.

ADDITIONAL BOOKS RECOMMENDED

1. Woodruff P. David, “Sketching as a Tool for Numerical Linear Algebra”, Foundations and Trends® in Theoretical Computer Science, now Publishers, USA.

Course Outcomes

At the end of the course, students will

CO1	have knowledge for types of Big Data, Design goals of Big Data platforms, and where in the systems landscape these platforms fall.
CO2	have information about distributed programming models for Big Data, including Map Reduce, Stream processing and Graph processing.
CO3	have learned runtime Systems for Big Data platforms and their optimizations on commodity clusters and Clouds.
CO4	be familiar with scaling data Science algorithms and analytics using Big Data platforms.
CO5	be able to configure, use different data mining software tools and develop applications to achieve scalable systems.

M. Tech. – I Semester – II	L	T	P	C
CSEDS606: ARTIFICIAL INTELLIGENCE (CORE ELECTIVE-2/3/4)	3	0	2	4

Course Objective	
1	To introduce the basic concepts of Artificial Intelligence (AI), with illustrations of current state of the art research, tools and applications.
2	To understand the basic areas of AI including problem solving, knowledge representation, heuristic, reasoning, decision making, planning and statistical methods.
3	To identify the type of an AI problem and apply it for search inference, decision making under uncertainty, game theory etc.
4	To describe the knowledge representation techniques, strengths and limitations of various state-space search algorithms, and choose the appropriate algorithm.
5	To introduce advanced topics of AI such as planning, Bayes networks, natural language processing and Expert systems.

INTRODUCTION TO AI AND INTELLIGENT AGENTS	(03 Hours)
Basic concepts of Intelligence, Scope and View of AI, Applications of AI, Turing Test, Intelligent Behavior, Intelligent Agents, AI Techniques, AI-Problem formulation, AI Applications, Production Systems, Control Strategies.	
PROBLEM SOLVING	**(08 Hours)**
Defining the problems as a State Space Search and Production Systems, Production Characteristics, Production System Characteristics, And issues in the Design of Search Programs, Additional Problems. Informed and uninformed search strategies: Generate-And-Test, Breadth first search, Depth first search, Hill climbing, Best first search, A* algorithm, AO* Algorithm, Iterative Deepening Search, IDA*, Recursive Best First Search, Constraint propagation, Neural, Stochastic, and Evolutionary search algorithms, Constraint Satisfaction and Heuristic Repair, Applications.	
KNOWLEDGE REPRESENTATION AND REASONING	**(06 Hours)**
Knowledge representation - Production based system, Frame based system, Knowledge representation using Predicate logic, Introduction to predicate calculus, Rule based representations, Declarative / Logical formalisms, Knowledge bases and Inference, Reasoning in uncertain environments, Logic-Structured based Knowledge representation, Inference – Backward chaining, Forward chaining, Rule value approach, Fuzzy reasoning – Certainty factors, Bayesian Theory-Bayesian Network-Dempster – Shafer theory, Symbolic Logic under Uncertainty : Non-monotonic Reasoning, Logics for non-monotonic reasoning, Statistical Reasoning :	

Probability and Bayes Theorem, Certainty factors, Probabilistic Graphical Models, Bayesian Networks, Markov Networks.	
GAME PLAYING AND PLANNING	**(06 Hours)**
Introduction, Example Domain: Overview, MiniMax, Alpha-Beta Cut-off, Refinements, Iterative deepening, The Blocks World, Components of a Planning System, Goal Stack Planning, Nonlinear Planning Using Constraint Posting, Hierarchical Planning, Reactive Systems, Other Planning Techniques, Recent applications.	
MULTI GAME THEORY	**(02 Hours)**
Introduction, Behavioral game theory: Dictator, Ultimatum and trust games, Mixed strategy equilibrium, Bargaining, Dominant solvable games, Coordination games, Signaling and reputation, Types of learning Reinforcement, Belief, Imitation, Stochastic game theory, Evolutionary games and Markov games for multi-agent reinforcement learning, Economic Reasoning and Artificial Intelligence, Designing games: Cooperative games, Voting, Auctions, Elicitation, Scoring rules, Decision Making and Utility Theory, Adaptive decision making, Analyzing games: Combinatorial games, Zero-sum games, General-sum games, Nash Equilibrium, Correlated Equilibrium, Price of anarchy.	
EXPERT SYSTEMS	**(04 Hours)**
Expert Systems – Architecture of Expert Systems, Roles of Expert Systems – Knowledge Acquisition – Meta Knowledge, Heuristics, Typical Expert Systems – MYCIN, DART, XOON, Expert Systems Shells.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Introduction to PROLOG programming.
2	Implement Informed and uniformed based search techniques.
3	Implement various algorithms based on game theory.
4	Practical based on fuzzy logic-based application.
5	Practical based on statistical methods.
6	Implement an expert system for real applications.
7	Practical based on multilayer perceptron.
8	Implement neural network-based application

BOOKS RECOMMENDED (LATEST EDITION)

1. Stuart Russell and Peter Norvig, "Artificial Intelligence: A Modern Approach", Prentice-Hall.
2. Nils J. Nilsson, "Artificial Intelligence: A New Synthesis", Morgan-Kaufmann.
3. Elaine Rich and Kevin Knight, "Artificial Intelligence", Tata McGraw-Hill.
4. W. Patterson, 'Introduction to Artificial Intelligence and Expert Systems', Prentice Hall of India.
5. I. Bratko, "Prolog Programming for Artificial Intelligence", Addison-Wesley.

ADDITIONAL BOOKS RECOMMENDED

1. Donald A.Waterman, "A Guide to Expert Systems", Pearson Education.
2. David Poole, Alan Mackworth, Artificial Intelligence: Foundations for Computational Agents, Cambridge Univ. Press.
3. J. Han and M. Kamber, Mining: Data Concepts and Techniques, 3rd Edition, Morgan Kaufman.
4. Hastie, Tibshirani, Friedman, "The elements of statistical learning", second edition, Springer.

Course Outcomes **At the end of the course, students will**	
CO1	be able to understand foundational principles, mathematical tools, program paradigms and fundamental issues, challenges of artificial intelligence, formal methods of knowledge representation, logic and reasoning.
CO2	be able to apply intelligent agents for artificial intelligence programming techniques, Fuzzy logic for problem solving and semantic rules for reasoning and inference to real world problems.
CO3	be able to analyze and formalize the problem as a state space, graph, design heuristics and select amongst different search or game-based techniques to solve them.
CO4	be able to evaluate the performance of an informed and uninformed search strategies, fuzzy logic, and expert system and connectionist models based systems.
CO5	be able to design the application on different artificial intelligence techniques like heuristic, game search algorithms, fuzzy, expert system and neural network.

M. Tech. – I Semester – II	**L**	**T**	**P**	**C**
CSEDS608: DATA MINING AND DATA WAREHOUSING (CORE ELECTIVE-2/3/4)	**3**	**0**	**2**	**4**

Course Objective	
1	To introduce students to the basic concepts and techniques of Data Mining.
2	To introduce a wide range of association, clustering, estimation, prediction, and classification algorithms.
3	To introduce mathematical statistics foundations of the Data Mining Algorithms.
4	To introduce basic principles, concepts and applications of Data Warehousing.
5	To build a data mining application from a data warehouse to solve real problems.

OVERVIEW	**(04 Hours)**
Introduction, Data Mining Issues, Data Mining Metrics, Data Mining from a Database Perspective, Data Mining Techniques: Classification, Statistical-Based Algorithms, Decision Tree -Based Algorithms, Neural Network-Based Algorithms, Rule-Based Algorithms, Combining Techniques; Similarity and Distance Measures, Hierarchical Algorithms, Partitioned Algorithms, Clustering Large Databases, Clustering with Categorical Attributes; Basic Algorithms, Advanced Association Rule Techniques, Measuring the Quality of Rules	
MINING STREAM, TIME SERIES AND SEQUENCE DATA	**(10 Hours)**
Mining Data Streams, Methodologies for Stream Data Processing and Stream Data Systems, Frequent-Pattern Mining in Data Streams, Classification of Dynamic Data Streams, Clustering Evolving Data Streams; Trend Analysis, Similarity Search in Time Series Analysis, Sequential Pattern Mining in Transactional Databases, Constraint-Based Mining of Sequential Patterns, Periodicity Analysis for Time-Related Sequence Data; Mining Sequence Patterns, Alignment of Sequences, Hidden Markov Model for Sequence Analysis.	
MULTIMEDIA DATA MINING	**(08 Hours)**
Multimedia Data, Similarity Search in Multimedia Data, Multidimensional Analysis of Multimedia Data, Classification and Prediction Analysis of Multimedia Data, Mining Associations in Multimedia Data, Audio and Video Data Mining.	
SPATIAL DATA MINING	**(08 Hours)**
Spatial Data, Mining Spatial Association and Co-location Patterns, Spatial Classification and Spatial Trend Analysis, Spatial Clustering Methods, Mining Raster Databases	
DATA WAREHOUSING	**(06 Hours)**

Review of Data Warehouse, Multidimensional Data Model, Data Cubes, Process Architecture, OLAP Operations, Stream OLAP and Stream Data Cubes, Generalization of Structured Data, Aggregation and Approximation in Spatial and Multimedia Data Generalization, Generalization of Class Composition Hierarchies, Construction and Mining of Object Cubes, Generalization-Based Mining of Plan Databases by Divide-and-Conquer, Spatial Data Cube Construction and Spatial OLAP.	
APPLICATIONS AND OTHER DM TECHNIQUES	**(06 Hours)**
Mining Event Sequences, Visual DM, Data Stream Mining, Multimedia Mining, Spatial Mining.	
Practical assignment will be based on the coverage of the above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Implementation of an application of a KDD process.
2	Analysis of Data Mining Techniques with Implementations using Java, Python etc.
3	Implementation of Nearest Neighbor Learning and Decision Trees.
4	Analysis of Splitting and Merging Clusters.
5	Implementation of association rule mining algorithms.
6	Mini Project: Implementation of Selected Journal Papers.

BOOKS RECOMMENDED (LATEST EDITION)

1. Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufman.
2. Ville, "Decision Trees for Business Intelligence and Data Mining: Using SAS Enterprise Miner", SAS.
3. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining", Addison Wesley.
4. Tom Soukup, Ian Davidson, "Visual Data Mining: Techniques and Tools for Data Visualization and Mining", Wiley.
5. Alex Berson, Stephen J. Smith, "Data Warehousing, Data Mining, and OLAP", MGH.

Course Outcomes **At the end of the course, students will**	
CO1	be able to identify the key processes of data mining, data warehousing and knowledge discovery process and understand the basic principles and algorithms used in practical data mining.
CO2	be able to apply data mining techniques to solve problems in other disciplines in a mathematical way.
CO3	be able to analyze the algorithms used in practical data mining and their strengths and weaknesses.
CO4	be able to evaluate different strategies of data warehousing techniques and data mining algorithms.

CO5	be able to design data mining algorithms for real time applications.

M. Tech. – I Semester – II	L	T	P	C
CSEDS610: NATURAL LANGUAGE PROCESSING (CORE ELECTIVE-2/3/4)	**3**	**0**	**2**	**4**

Course Objective	
1	To comprehend natural language processing in order to extract information.
2	To understand information about language-specific tasks and learning models.
3	To investigate the use of artificial intelligence to comprehend the semantics of text data.
4	To know about text processing at syntactic, semantic, and pragmatic levels.
5	To understand data extraction from unstructured text by identifying references to named entities as well as stated relationships between such entities.

INTRODUCTION AND LANGUAGE MODELING	**(12 Hours)**
Introduction to Computational Linguistics, Word Meaning, Distributional Semantics, Word Sense Disambiguation, Sequence Models, N-gram Language Models, Feed forward Neural Language Models, Word Embedding, Recurrent Neural Language Models, Tokenization, Lemmatization, Stemming, Sentence Segmentation, POS Tagging and Sequence Labeling, Structured Perceptron, Viterbi – Loss, Augmented Structured Prediction, Neural Text Models and Tasks.	
INFORMATION EXTRACTION	**(10 Hours)**
Information Extraction from Text, Sequential Labeling, Named Entity Recognition, Semantic Lexicon Induction, Relation Extraction, Paraphrases Inference Rules, Summarization, Event Extraction, Opinion Extraction, Temporal Information Extraction, Open Information Extraction, Knowledge based Population, Narrative Event Chains and Script Learning, Knowledge Graph Augmented Neural Networks for Natural Language.	
MACHINE TRANSLATION AND ENCODER-DECODER MODELS	**(10 Hours)**
Machine Translation, Encoder-Decoder Models, Beam Search, Attention Models, Multilingual Models, Syntax, Trees, Parsing, Transition based Dependency Parsing, Graph based Dependency Parsing, Transfer Learning, Deep Generative Models for Natural Language Data, Text Analytics, Text Mining, Information Extraction with AQL-Conversational AI.	
APPLICATION AND CASE STUDIES	**(10 Hours)**
Application: Spelling Correction, Sentiment Analysis, Word Sense Disambiguation, Text Classification, Machine Translation, Question Answering System, Intent Detection, False Fact Detection .	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Create an application in Python with the NLTK library to tokenize the words present in a paragraph.
2	Perform tasks with NLTK (Natural Language Toolkit).
3	Tasks to be Performed in SpacCy Library.
4	Practicals based on huggingface library.
5	Text Classification using movie reviews database, etc.
6	Practical implementation of application and case study.

BOOKS RECOMMENDED (LATEST EDITION)

1. Emily Bender, “Linguistics Fundamentals for NLP”, Morgan Claypool Publishers.
2. Jacob Eisenstein, “Natural Language Processing”, The MIT Press.
3. Dan Jurafsky, James H. Martin, “Speech and Language Processing”, Prentice Hall.
4. Chris Manning, HinrichSchutze, “Foundations of Statistical Natural Language Processing”, The MIT Press.
5. Pushpak Bhattacharyya, “Machine Translation”, CRC Press.

Course Outcomes At the end of the course, students will	
CO1	be able to understand how language works, including the word structure, sentence structure, and meaning.
CO2	be able to learn how to reframe NLP problems as learning and inference tasks, as well as how to deal with the associated computational challenges
CO3	be able to use text processing at the syntactic, semantic, and pragmatic levels.
CO4	be able to learn about text mining and manipulation techniques.
CO5	be able to retrieve information from the text and can use it for decision making.

M. Tech. – I Semester – II	L	T	P	C
CSEDS612: DATA SCIENCE FOR SOFTWARE ENGINEERING (CORE ELECTIVE-2/3/4)	**3**	**0**	2	**4**

Course Objective	
1	To understand various tools of Software Engineering.
2	To understand the capability of software engineering principles to analyze data science applications to make appropriate decisions.
3	To learn various methods and principles of software engineering for data science applications.
4	To learn integration of software engineering principles with data science applications.
5	To learn how to use software engineering for data science.

FORMAL SOFTWARE ENGINEERING	**(06 Hours)**
Formal specifications, Techniques, Verification and Validation, Theorem Provers, Model checking, modeling concurrent systems, Temporal logics, CTL & LTL and model checking, SAT Solvers, Testing Techniques, Test Case Generation	
SOFTWARE REQUIREMENTS AND ESTIMATION	**(04 Hours)**
Software Requirements: What and Why, Software Requirements Engineering, Software Requirements Management, Software Requirements Modeling, Software Estimation, Size Estimation, Effort, Schedule and Cost Estimation, Tools for Requirements Management and Estimation.	
SOFTWARE DEVELOPMENT METHODOLOGIES	**(04 Hours)**
Introduction to Software Engineering, A Generic View of Process, Process Models, Software Requirements, Design Engineering, Creating an Architectural Design, Modeling Component.	
SOFTWARE PROCESS AND PROJECT MANAGEMENT	**(04 Hours)**
Software Process Maturity, Process Reference Models, Software Project Management Renaissance, Life-Cycle Phases and Process artifacts, Workflows and Checkpoints of Process, Process Planning, Project Organizations, Project Control and Process Instrumentation, CCPDS-R Case Study and Future Software Project Management Practices.	
FUNDAMENTALS OF OBJECT ORIENTED DESIGN IN UML	**(04 Hours)**
Static and Dynamic Models, Necessity of Modeling, UML Diagrams, Class Diagrams, Interaction Diagrams, Collaboration Diagram, Sequence Diagram, State Chart Diagram, Activity Diagram, Implementation Diagram.	

USER INTERFACE	**(04 Hours)**
Module Introduction, Objectives of Usability, How to Approach Usability, Designing with Usability in mind, Measuring Usability, Guidelines for User Interface Design, User Interface Elements.	
SOFTWARE QUALITY ASSURANCE AND TESTING	**(04 Hours)**
Software Quality Assurance and Standards, Quality Standards, Software Testing Strategy and Environment, Building Software Testing Process, Software Testing Techniques, Software Testing Tools, Testing Process-Seven Step Testing Process, Specialized Testing Responsibilities.	
DATA SCIENCE PERSPECTIVE FOR SOFTWARE ENGINEERING	**(12 Hours)**
Diverse Sets of Data, Category of Data, Combining Quantitative and Qualitative Methods, Structuring and Summarizing Unstructured Software Data, Validate and Calibrate Data, Generation of Requirement Specifications, Automatic Code Documentation; Software Project Cost Estimation, Software Quality Prediction, Semi-Automatic Refactoring, Prioritization, Automatic Bug Assignment and Test Cases Generation; Case Study-Search Engine: Working of Search Engine, Content Quality Strategy, Control Crawling, Indexing and Ranking, Search Appearance, Optimization.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Working with software engineering software SPIN.
2	Working with a variety of modules for software engineering.
3	Working with testing of the software project.
4	To develop the software engineering prototype of the application.
5	To analyze the software using a model checker.

BOOKS RECOMMENDED (LATEST EDITION)

1. Roger S. Pressman, “Software Engineering: A Practitioner's Approach”, McGraw Hill Higher Education.
2. Ian Sommerville, “Software Engineering”, Pearson Education.
3. Carlo Ghezzi, Mehdi Jazayeri, Dino Mandrioli, “Fundamentals of Software Engineering”, Pearson.
4. Hans van Vliet, “Software Engineering: Principles and Practice”, Wiley.
5. Tim Menzies, Laurie Williams, Thomas Zimmermann, “Perspectives on Data Science for Software Engineering”.

Course Outcomes **At the end of the course, students will**	
CO1	have knowledge about software engineering tools for integrated development environments, syntax checking, testing, debugging, and version control.
CO2	be able to apply software engineering principles to solve Data Science applications.
CO3	be able to critically analyze the Data Science problems to apply software engineering solutions.
CO4	be able to evaluate various Data Science applications using software engineering principles.
CO5	be able to design software engineering principles based applications using Data Science principles.

M. Tech. – I Semester – II	L	T	P	C
CSEDS614: BIG DATA ANALYTICS AND LARGE-SCALE COMPUTING (CORE ELECTIVE-2/3/4)	**3**	**0**	**2**	**4**

Course Objective	
1	To learn the basics of big data, its characteristics, big data management issues, processing and applications with the help of big data platforms and storage models for big data management.
2	To learn the management and analysis of big data using technology like Hadoop, NoSql, MapReduce, PIG & HIVE.
3	To apply the data mining algorithms on big data for scalability of the real time applications.
4	To develop research interest towards advances in data mining by analyzing the available approaches with the help of evaluating parameters.
5	To build big data analytics and management systems with visualization using the latest technology to solve real problems.

INTRODUCTION	**(04 Hours)**
Definition of Big Data, Source of Big Data, Convergence of Key Trends, Unstructured Data, Industry Examples of Big Data, Web Analytics, Fraud and Risk Associated with Big Data, Credit Risk Management, Big Data in Algorithmic Trading, Healthcare, Medicine, Marketing and Advertising, Big Data Technologies, Introduction to Hadoop and Spark, Open Source Technologies, Cloud, Mobile Business Intelligence, Crowd Sourcing Analytics, Inter and Trans Firewall Analytics.	
BIG DATA ANALYTICS	**(06 Hours)**
Big Data Processing: Batch Data Processing and Stream Data Processing, Computing Environments for Big Data Analytics, Implementation of Batch and Real Time Event Processing: Integration of Disparate Data Stores/Data Lake, Mapping Data to the Programming Framework, Connecting and Extracting Data from Storage, Transforming Data for Processing, Querying.	
DISTRIBUTED FILE SYSTEM HADOOP	**(08 Hours)**
Introduction, HDFS Daemons, Different Methods to HDFS Access, Hadoop, Features, Google File System Features, Phases involved in Map Reduce, Architecture, Execution of MapReduce Jobs, Monitoring the progress of job flows, Building Blocks of Hadoop MapReduce. Data format, Analyzing data with Hadoop, Scaling Out, Hadoop Streaming, Hadoop Pipes, Design of Hadoop Distributed File System, MapReduce, HDFS Concepts: Java Interface, Data Flow, Hadoop I/O, Data integrity, Compression, Serialization, Avro, File-based Data Structures, Mahout, Pig, Hive, HBase.	

DISTRIBUTED MACHINE LEARNING	**(08 Hours)**
Review of Machine Learning: Supervised and Unsupervised Learning, Linear algebra; Classification Formulation, Closed Form Solution, Computational Complexity, Grid Search, Computation Storage Communication, Probabilistic Prediction, Backpropagation Graph and Compute Gradients for Model Training, Automatic Differentiation Graph-Level Optimization Parallelization/Distributed Training Data Layout and Distributed Linear Regression and Distributed Logistic Regression, Placement Kernel Optimizations, Memory Optimizations, Distributed Principal Component Analysis, Regularization and Optimization for Training Deep Neural Networks, Sequence Modeling, Federated Learning.	
BIG DATA ANALYSIS WITH MLLIB, SPARKSQL AND GRAPHX	**(05 Hours)**
HBase, Data Model and Implementations, HBase Clients, HBase Examples, Praxis, Cassandra, Cassandra data Model, Cassandra Examples, Cassandra Clients, Hadoop Integration, Hive, Data Types and File Formats, HiveQL Data Definition, HiveQL Data Manipulation, HiveQL Queries, Applications on Big Data Using Pig and Hive, Data Processing Operators in Pig, Fundamentals of ZooKeeper, K-Means Clustering, Decision Trees, Random Forests, Recommenders, Table in Spark, Higher Level Declarative Programming, Network Structure, Computing Graph Statistics.	
BIG DATA STORAGE MODELS	**(06 Hours)**
Introduction, NoSQL Databases, Need, Types, Comparison with RDBMS, Architecture and Features of NoSQL Databases: Distributed Hash-table, Key-Value Storage Model, Document Storage Model, Graph Storage Models, Lambda Architecture, Data Ingestion, Design and Provision Compute Resources, Storage Technology, Streaming Units, Configuration of Clusters for Latency and Throughput, Output Visualization.	
SCALABLE ALGORITHMS	**(05 Hours)**
Mining Big Data, Centrality, Similarity, Al-Distances Sketches, Community Detection, Link Analysis, Spectral Techniques, MapReduce, Pig Latin, and NoSQL, Algorithms for Detecting Similar Items, Recommendation Systems, Data Stream Analysis Algorithms, Detecting Frequent Items, Data Ingestion, Storage of Data, Data Transfer, Compute Clusters and Configuration of Design.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Working with various functions of Hadoop MapReduce.
2	Working with pySpark and RDDs.

3	Regression and classification in Spark.
4	Data analysis with PCA in Spark.
5	Hands-on with MLlib and SparkSQL.
6	Use cases and implementation for Big data management and large scale machine learning algorithms.

BOOKS RECOMMENDED (LATEST EDITION)

1. Ron Bekkerman, Mikhail Bilenko, John Langford, “Scaling up Machine Learning: Parallel and Distributed Approaches”, Cambridge University Press.
2. Michael Minelli, Michele Chambers, Ambiga Dhiraj, “Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley.
3. Michael Berthold, David J. Hand, “Intelligent Data Analysis”, Springer.
4. Tom White, “Hadoop: The Definitive Guide”, O’reilly Media.
5. Arshdeep Bahga, Vijay Madisetti, “Big Data Science & Analytics: A Hands on Approach “, VPT.

ADDITIONAL BOOKS RECOMMENDED

1. Edward Capriolo, Dean Wampler, and Jason Rutherglen, “Programming Hive”, O'Reilly.
2. Lars George, “HBase: The Definitive Guide”, O'Reilly.
3. Eben Hewitt, “Cassandra: The Definitive Guide”, O'Reilly.
4. Alan Gates, “Programming Pig”, O'Reilly.
5. Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills, “Advanced Analytics with Spark”, O'Reilly.
6. Holden Karau, Andy Konwinski, Patrick Wendell, and MateiZaharia,Learning Spark, O'Reilly.
7. Jure Leskovec, Stanford Univ.AnandRajaraman, Milliway Labs, Jeffrey D. Ullman, “Mining of Massive Datasets”, Cambridge University Press.
8. Ron Bekkerman, Mikhail Bilenko and John Langford, “Scaling up Machine Learning: Parallel and Distributed Approaches”, Cambridge University Press.
9. Arvind Sathi, “Big Data Analytics: Disruptive Technologies for Changing the Game”, MC Press.
10. Tom Plunkett, Brian Macdonald et al, “Oracle Big Data Handbook”, Oracle Press.
11. Jay Liebowitz, “Big Data and Business analytics”, CRC press.

Course Outcomes

At the end of the course, students will

CO1	have knowledge of the key issues in big data management and its associated applications in intelligent business and scientific computing.

CO2	be able to apply theoretical foundations of mining algorithms for the usage applicability of business, engineering and scientific problems for big data processing and scalability.
CO3	be able to analyze Hadoop related tools such as HBase, Cassandra, and Hive for big data analytics.
CO4	be able to evaluate the big data analytics applications and evaluation measures to have a productive solution.
CO5	be able to build a complete business data analytics solution for any real time problem.

M. Tech. – I Semester – II	L	T	P	C
CSEDS616: CYBER PHYSICAL SYSTEMS (CORE ELECTIVE-2/3/4)	**3**	**0**	**2**	**4**

Course Objective	
1	The students will have an understanding of the cyber physical systems and the corresponding important research challenges in this area.
2	To learn the current state of art in CPS domain and the details regarding necessary principles required for future CPS.
3	A third objective is improving critical reading, presentation, and research skills.

Course Syllabus:
Introduction to Cyber-Physical Systems. The Industrial Revolution 4.0. Motivation for the IR 4.0. Why are the CPS touted as IR 4.0? Cyber-Physical Systems (CPS) in the real world.
Basic principles of design and validation of CPS. Basic characteristics of the CPSs. The Internet of Things. The Industrial Internet of Things. The Wireless Sensor Networks and the RFID devices as the actors of the CPSs. The Ubiquitous and the Pervasive Computing paradigm introduced by the CPSs. The Applications of the Wireless Sensor Networks. The role of the Internet of Things in realizing Smart Applications. The Characteristics and the issues of deployment.
The CPS Hardware Platforms: Processors. Types of Processor. The Processors Design issues. Parallelism. Embedded Processors. Harvard Architecture: Pros and Cons. The Sensors and Actuators. Models of Sensors and Actuators. Common Sensors. Actuators. Memory Architectures. Memory Technologies. Memory Hierarchy. Memory Models. Types of memory in the CPSs. Input and Output Hardware. The design issues. The Analog to Digital convertor.
The Real time Operating Systems for the WSN devices. Characteristics. Issues. Thread Scheduling. Basics of Scheduling. Rate Monotonic Scheduling. The Earliest Deadline First Scheduling. Scheduling and Mutual Exclusion. Multiprocessor Scheduling. Sequential Software in a Concurrent World. Multitasking. Imperative Programs. Case studies of the typical OSs. TinyOS, nesC and Contiki. The Simulators for the WSN devices. The CPS Network - WirelessHart, CAN, Automotive Ethernet.
Formal Methods for Safety Assurance of Cyber-Physical Systems: Advanced Automata based modelling and analysis, Basic introduction and examples, Timed and Hybrid Automata, Definition of trajectories, Formal Analysis: Flow pipe construction, reachability analysis. Analysis of CPS Software: Weakest Pre-conditions, Bounded Model checking, CPS software verification: Frama-C, CBMC

Secure Deployment of CPS: Attack models, Secure Task mapping and Partitioning, State estimation for attack detection Automotive Case study: Vehicle ABS hacking Power Distribution Case study: Attacks on SmartGrids.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be based on the content discussed in class.)

BOOKS RECOMMENDED (LATEST EDITION)

1. E. A. Lee and S. A. Seshia, Introduction to Embedded Systems - A Cyber-Physical Systems Approach, The MIT Press.
2. Rajeev Alur, Principles of Cyber-Physical Systems, The MIT Press.
3. Zeadally S. and NafaâJabeur, Cyber Physical System Design With Senor Networking Technologies, The IET Press.

Course Outcomes

At the end of the course, students will be able to

CO1	Define embedded systems and cyber-physical systems (CPS).
CO2	Understand the different paradigms of computing and how the ubiquitous and pervasive computing affects the Cyber physical systems.
CO3	Analyze the design issues associated with different hardware functional units of the CPSs.
CO4	Analyze the performance impact of thread scheduling algorithms in the CPSs.
CO5	Understand various modelling formalisms for CPS, viz. hybrid automata, timed automata, state-space methods and the likes.

M. Tech. – I Semester – II	L	T	P	C
CSEDS618: MACHINE LEARNING FOR SECURITY(CORE ELECTIVE-2/3/4)	**3**	**0**	**2**	**4**

Course Objectives	
1	To DESCRIBE the fundamental concepts of machine learning for devising security mechanisms.
2	To ENUMERATE the techniques for Intrusion Detection and Malware detection and analysis using Machine Learning.
3	To learn the machine learning tecniques for network traffic analysis
4	To analyse the machine learning approaches for security for probable abuse by the adversary.
5	To design secure machine learning based schemes for malware detection and intrusion detection.

INTRODUCTION & REVIEW OF THE MACHINE LEARNING BASICS	**(02 Hours)**
Review of the basic concepts in Linear Algebra, Probability and Statistics. Introduction to the ML techniques. Machine Learning problems viz. Classification, Regression, Clustering, Association rule learning, Structured output, Ranking. The Supervised and Unsupervised learning algorithms. Linear Regression, Gradient descent for convex functions, Logistics Regression and Bayesian Classification Support Vector Machines, Decision Tree and Random Forest, Neural Networks, DNNs , Ensemble learning. Principal Components Analysis. Un-supervised learning algorithms: K-means for clustering problems, K-NN (k nearest neighbors). Apriori algorithm for association rule learning problems. Generative vs Discriminative learning. Empirical Risk Minimization, loss functions, VC dimension. Data partitioning (Train/test/Validation), cross-validation, Biases and Variances, Regularization.	
MACHINE LEARNING FOR SECURITY	**(04 Hours)**
Introduction to Information Assurance. Review of Cybersecurity Solutions: Proactive Security Solutions, Reactive Security Solutions: Misuse/Signature Detection, Anomaly Detection, Hybrid Detection, Scan Detection. Profiling Modules. Understanding the Fundamental Problems of Machine-Learning Methods in Cybersecurity. Incremental Learning in Cyber infrastructures. Feature Selection/Extraction for Data with Evolving Characteristics. Privacy-Preserving Data Mining. Motivation for ML in security with real-world case studies. Topics of interest in applications of machine learning for security.	
MACHINE LEARNING TECHNQIUES FOR INTRUSION DETECTION	**(08 Hours)**
Emerging Challenges in Cyber Security for Intrusion Detection: Unifying the Current Anomaly Detection Systems, Network Traffic Anomaly Detection. Imbalanced Learning Problem and Advanced Evaluation Metrics for IDS. Reliable Evaluation Data Sets or Data Generation Tools. Privacy Issues in Network Anomaly	

Detection. Machine Learning Techniques: for Anomaly Detection, for Misuse/Signature detection, for Hybrid detection, for Scan detection.Cost-Sensitive Modeling for Intrusion Detection. Data Cleaning and Enriched Representations for Anomaly Detection in System Calls.	
MACHINE LEARNING TECHNQIUES FOR MALWARE ANALYSIS	**(08 Hours)**
Emerging Cyber Threats in malwares: Threats from Malware, Botnets, Cyber Warfare, Mobile Communication. Cyber Crimes. Malware Analysis: Feature generation, Features to Classification. Taxonomy of malware analysis approaches based on machine learning. Malware Detection, Similarity Analysis, Category Detection. Feature Extraction. PE Features. Supervised, Unsupervised and Semi-supervised learning algorithms for Malware Detection. Using Deep Learning Approaches: Generative Adversarial Networks.	
NETWORK TRAFFIC ANALYSIS & WEB ABUSE DETECTION	**(08 Hours)**
Machine Learning for Profiling Network Traffic: Theory of Network defense (access control, authentication, detecting in-network attackers, data-centric security, honeypots), Predictive model for classifying network attacks.	
MACHINE LEARNING IN PRIVACY PRESERVATION	**(06 Hours)**
k-anonymity; l-diversity; deferentially private data storage/release; verifiable differential privacy; privacy-preserving inference of social networking data; privacy-preserving recommender system; privacy versus utility. Machine learning techniques for Privacy Preserving Data Mining.	
ADVERSARIAL MACHINE LEARNING	**(06 Hours)**
Adversarial Machine Learning: Motivation and Background. Practical Scenarios and Examples. Modelling the Adversary: Attack Surface Adversary Goals Adversary capabilities. Taxonomy of Adversarial Attacks on Machine Learning: Influence Specificity Security Violation. Data poisoning; Perturbation; Defense mechanism; Generative Adversarial Networks. A peep into Industry Perspectives: Theme of inference Secure Software Development Life Cycle or Secure Development Cycle. Key Inferences in terms of Security gaps, Suggested panacea.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)	

BOOKS RECOMMENDED (LATEST EDITION)

1. Clarence Chio, David Freeman, Machine Learning and Security. Protecting Systems with Data and Algorithms, O'Reilly Media Publications.
2. Marcus A. Maloof (Ed.), Machine Learning and Data Mining for Computer Security: Methods and Applications, Springer-Verlag London Limited.

3. Sumeet Dua and Xian Du, Data Mining and Machine Learning in Cybersecurity. CRC Press, Taylor and Francis Group, LLC.
4. Research Papers Prescribed in the class.

Course Outcomes **At the end of the course, students will**	
CO1	have a knowledge of the limitations of the conventional security software in the wake of machine learning based attacks on the security software
CO2	be able to apply the concepts machine learning based intrusion detection to analyze the IDSs.
CO3	be able to analyze the malware analysis and mitigation based solutions for the probable threats therein.
CO4	be able to design the threat models based on machine learning approaches for network analysis.
CO5	be able to use the concepts of machine learning to prevent security design faults.

M. Tech. – I Semester – II	**L**	**T**	**P**	**C**
CSEDS620: BUSINESS DATA ANALYTICS (INSTITUTE ELECTIVE)	**3**	**0**	**2**	**4**

Course Objective	
1	Gaining fundamental Knowledge of Business Analytics and Data Science.
2	To become acquainted with the procedures needed to develop, report, and analyze professional data.
3	To deepen analytical skills and investigate data to establish new relationships and patterns.
4	To optimize business decisions and create competitive advantage with Data analytics.
5	To recognize the importance of Visualization tools for Data Analytics in Business.

INTRODUCTION	**(06 Hours)**
Introduction to Business Analytics, Applications, Components, Types of Business Analytics, Transaction Processing versus Analytic Processing, Big Data and Its Components.	
DATA WAREHOUSE	**(12 Hours)**
Sources of Data, Organization of Data, Types of Data (Raw and Processed), Introduction to Data Warehouse, Multidimensional Data Model, Data Marts, Data Integration, ELT, Concepts of OLAP and Data Cube, OLAP Operations, Dimensional Data Modeling - Star, Snowflake Schemas, Hierarchies, Aggregations.	
VISUALIZING DATA	**(08 Hours)**
Structure of Visualization, Organization of Data, Importance of Data Quality, Dealing with Missing and Incomplete Data, Data Classification, Different Kinds of Plots, Charts and Their Usage, Dashboard and Interactive Plots, Visual Data Analysis Techniques, Interaction Techniques, Creating Animated Visualizations.	
DATA MINING FOR BUSINESS	**(10 Hours)**
Data Mining Process, Data Mining Algorithms (Supervised and Unsupervised), Definition and Concept of Data Mining, Benefits of Data Mining, Data Mining Tasks, Text Mining, Web Mining, Spatial Mining, Process Mining, Social Media Analytics, Social Media Metrics.	
APPLICATIONS OF DATA ANALYTICS IN BUSINESS	**(06 Hours)**
Application of Business Analysis using Tableau, BI Tools: IT analytics, Retail Analytics, Process Analytics, Financial Analytics, Healthcare Analytics, Supply Chain Analytics.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be changed every year and will be notified on the website.)	
1	Working with R studio software to use various data types and objects.
2	Working with Tableau, Data transformation with Visual concepts.
3	Working Power BI with Power Apps and Power Automate to build business applications and automate workflows.
4	Working with Python Programming to solve data manipulation, analysis for business, etc.
5	Problems based on Data Mining techniques.

BOOKS RECOMMENDED (LATEST EDITION)

1. Ramesh Sharda, DursunDelen, Efraim Turban, and David King, “Business Intelligence, Analytics, and Data Science: A Managerial Perspective”, Pearson Education Limited.
2. Noah Iliinsky and Julie Steele, “Designing Data Visualizations”, O’Reilly.
3. Foster Provost and Tom Fawcett, “Data Science for Business: What You Need to Know”, O’Reilly.
4. Melissa Barker, Donald I. Barker, Nicholas F. Bormann, Debra Zahay, “Social Media Marketing: A Strategic Approach”, Cengage Learning.
5. GerKoole, “An Introduction to Business Analytics”, MG Books.

ADDITIONAL BOOKS RECOMMENDED

1. Laura Igual, Santi Seguí, “Introduction to Data Science”, Springer.
2. Michael Minelli, Michele Chambers, AmbigaDhiraj, “Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley.
3. ArshdeepBahga, Vijay Madisetti, “Big Data Science & Analytics: A Hands on Approach “, VPT.

Course Outcomes **At the end of the course, students will**	
CO1	have knowledge about various data tools and techniques needed in business decision making.
CO2	be able to apply different tools and functions of various software’s to visualize a variety of data in the appropriate form of visualization.
CO3	be able to critically analyze the business problems and apply analytical knowledge in big data.
CO4	be able to evaluate various data analytical techniques.
CO5	be able to design business analytical applications using Data Science principles for the decision making process.

M. Tech. – I Semester – II	L	T	P	C
CSEDS622: SOCIAL NETWORKS (INSTITUTE ELECTIVE)	**3**	**0**	**2**	**4**

Course Objective	
1	To understand the social network models, representation and analytics.
2	To identify the unique challenges involved in social network research.
3	To apply techniques for social network representation and analytics for real-word scenarios.
4	To analyse and evaluate the social network research solutions for real-world scenarios.

INTRODUCTION	**(08 Hours)**
Introduction to Social Networks, Networks as Information Maps, Networks as Conduits, Connections, Propinquity, Homophily	
SOCIAL NETWORK ANALYSIS	**(18 Hours)**
Mathematical Foundations, Data Collection, Data Management, Visualization, Centrality, Subgroups, Cliques, Clusters, Dyads and Triads, Density, Structural Holes, Weak Ties, Centrality, The Small World, Circles, and Communities, Multiplicity, Structural Similarity and Structural Equivalence	
SOCIAL NETWORKS AND DIFFUSION	**(08 Hours)**
Influence and Decision-Making, Epidemiology and Network Diffusion, Tipping Points and Thresholds	
Social Network Tools and Case Studies	**(08 Hours)**
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

List of Practical (Problem statements will be based on the content discussed in class.)

BOOKS RECOMMENDED (LATEST EDITION)

1. Borgatti SP, Everett MG, Johnson JC, “Analyzing Social Networks”, London, Sage Publication.
2. Kadushin C., “Understanding Social Networks: Theories, Concepts and Findings”, Oxford University Press.

3. Piet A.M. Kommers, Pedro Isaias, TomayessIssa, “Perspectives on Social Media: A Yearbook”, Taylor and Francis.
4. Newman Mark, “Networks: An Introduction”, Oxford university press.
5. Brath Richard, David Jonker, "Graph analysis and visualization: Discovering Business Opportunity in Linked Data", John Wiley & Sons.

Course Outcomes **At the end of the course, students will**	
CO1	have the knowledge of various social network representation, visualization and analytics tools and techniques.
CO2	be able to apply tools for social network data acquisition, management and analytics.
CO3	be able to analysethe social network research solutions for real-world scenarios.
CO4	be able to evaluate the different solutions for performance;
CO5	be able to design the social network analytics solution for the complex real-world problem.

M. Tech. – I Semester – II	L	T	P	C
CSEDS624: CYBER LAWS (INSTITUTE ELECTIVE)	**3**	**0**	**2**	**4**

Course Objective	
1	The course aims at acquainting the students with the basic concepts of Cyber Law and also puts those concepts in their practical perspective.
2	It also provides an elementary understanding of the authorities under IT Act as well as penalties and offences under IT Act.
3	It also covers overview of Intellectual Property Right and Trademark Related laws with respect to Cyber Space.
4	Student will get the knowledge about the E- Governance policies of India.

INTRODUCTION OF CYBER CRIMES & CYBER LAW	**(06 Hours)**
Understanding Cyber Crimes and Cyber Offences, Crime in context of Internet, Types of Crime in Internet, Crimes targeting Computers: Definition of Cyber Crime & Computer related Crimes, Constraint and Scope of Cyber Laws, Social Media and its Role in Cyber World, Fake News, Defamation, Online Advertising.	
PREVENTION OF CYBER CRIMES & IT ACT 2000	**(06 Hours)**
Prevention of Cyber Crimes & Frauds, Evolution of the IT Act 2000, Genesis and Necessity. Critical analysis & loop holes of The IT Act, 2000 in terms of cyber-crimes, Cyber Crimes: Freedom of speech in cyber space & human right issues.	
FEATURES OF IT ACT 2000 & AMENDMENTS	**(06 Hours)**
Salient features of the IT Act, 2000, Cyber Tribunal & Appellate Tribunal and other authorities under IT Act and their powers, Penalties & Offences under IT Act, Amendments under IT Act and Impact on other related Acts (Amendments): (a) Amendments to Indian Penal Code. (b) Amendments to Indian Evidence Act. (c) Amendments to Bankers Book Evidence Act. (d) Amendments to Reserve Bank of India Act.	
INDIAN PENAL LAW	**(06 Hours)**
Indian Penal Law and Cyber Crimes: (i) Fraud, (ii) Hacking, (iii) Mischief, Trespass (iv) Defamation (v) Stalking (vi) Spam, Issues of Internet Governance: (i) Freedom of Expression in Internet (ii) Issues of Censorship (iii) Hate Speech (iv) Sedition (v) Libel (vi) Subversion (vii) Privacy, Cyber Appellate Tribunal with Special Reference to the Cyber Regulation Appellate Tribunal (Procedures) Rules 2000.	
GLOBAL IT RULES & IPR	**(06 Hours)**
The Information Technology (Procedures and Safeguards for Interception, Monitoring and Decryption of	

Information) Rules, 2009 and Corresponding International Legislation in US, UK and Europe, The Information Technology (Procedures and Safeguards for Blocking the access of Information by Public) Rules, 2009 and Corresponding International Legislation in US, UK and Europe, The Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2009 and Corresponding International Legislation in US, UK and Europe, Intellectual Property Right (IPR).	
CYBER SPACE & E-GOVERNANCE IN INDIA	**(06 Hours)**
Cyber and Cyber Space with reference to Democracy and Sovereignty, Developments in Cyber law Jurisprudence, Role of law in Cyber World: Regulation of Cyber Space in India, Role of RBI and Legal Issues in case of e-commerce, E-Governance in India: Law, Policy, Practice.	
CYBER SPACE JURISDICTION	**(06 Hours)**
Cyber Space Jurisdiction (a) Jurisdiction issues under IT Act, 2000. (b) Traditional principals of Jurisdiction (c) Extra-terrestrial Jurisdiction (d) Case Laws on Cyber Space Jurisdiction (e) Taxation issues in Cyberspace.	
Practical assignments will be based on the coverage of above topics.	**(28 Hours)**
	(Total Contact Time: 42 Hours + 28 Hours = 70 Hours)

BOOKS RECOMMENDED (LATEST EDITION)

1. Vakul Sharma, "Information Technology Law and Practice - Cyber Laws and Laws Relating to E-Commerce", Universal Law Publishing - An imprint of LexisNexis.
2. Duggal Pavan, "Legal Framework on Electronic Commerce and Intellectual Property Rights in Cyberspace", Universal Law Publishing - An imprint of LexisNexis.
3. Santosh Kumar, "Cyber Laws & Cyber Crimes", WHITESMANN.
4. Yatindra Singh, "Cyber Laws: A Guide to Cyber Laws, Information Technology, Computer Software, Intellectual Property Rights, E-commerce, Taxation, Privacy, Etc. Along with Policies, Guidelines and Agreements", Universal Law Publishing.

Course Outcomes

At the end of the course, students will

CO1	Student will be able to understand the types of Crime in Internet, Crimes targeting Computers and Scope of Cyber Laws.
CO2	Student will be able to apply the cyber laws to related the various evidences of cybercrimes.
CO3	Student will be able to analyze the various evidences of cybercrimes to allied with the particular

	cyber law.
CO4	Student will be able to evaluate the particular intellectual property rights according to the cyber law.
CO5	Student will be able to design an application to counter the cybercrimes.

*