MACHINE LEARNING

Course Code: KG21CM601

L	T	P	C
3	0	0	3

B. Tech. III Year II - Semester

Prerequisites: Knowledge on Basics of Data Structures and statistical methods.

Course Objectives: The objectives of this course for the student are to:

1. Apply concept learning and supervised Algorithms on data samples.
2. Illustrate the Back Propagation Algorithm in Machine Learning.
3. Classify Bayesian, Computational Learning Theory and Instance Based Learning in Machine Learning.
4. Develop Genetic Algorithm, Learning Set of Rules and Reinforcement Learning approaches.
5. Discuss Inductive and Analytical Learning in Machine Learning.

Course Outcomes: After completion of this course, the students will be able to

CO1: Apply Concept Learning and supervised Algorithms on data samples.

CO2: Illustrate the Back Propagation Algorithm in Machine Learning.

CO3: Classify Bayesian, Computational Learning Theory and Instance Based Learning in Machine Learning.

CO4: Develop Genetic Algorithm, Learning Set of Rules and Reinforcement Learning approaches.

CO5: Discuss Inductive and Analytical Learning in Machine Learning.

UNIT-I

Introduction - Well-posed learning problems, designing a learning system, Perspectives and issues in machine learning, Concept learning and the general to specific ordering – introduction, a concept learning task, concept learning as search. Find-S: finding a maximally specific hypothesis, version spaces and the candidate elimination algorithm, remarks on version spaces and candidate elimination, inductive bias.

Decision Tree Learning: Introduction, decision tree representation, appropriate problems for decision tree learning, the basic decision tree learning algorithm, hypothesis space search in decision tree learning, inductive bias in decision tree learning, issues in decision tree learning.

UNIT-II

Artificial Neural Networks - 1: Introduction, neural network representation, appropriate problems for neural network learning, perceptions, multilayer networks and the back-propagation algorithm.

Artificial Neural Networks - 2: Remarks on the Back-Propagation algorithm, An illustrative example: face recognition, advanced topics inartificial neural networks.

Evaluation Hypotheses: Motivation, estimation hypothesis accuracy, basics of sampling theory, a general approach for deriving confidence intervals, difference in error of two hypotheses, comparing learning algorithms.

UNIT-III

Bayesian learning – Introduction, Baye's theorem, Baye's theorem and concept learning, Maximum Likelihood and least squared error hypotheses, maximum likelihood hypotheses for predicting probabilities, minimum

description length principle, Baye's optimal classifier, Gibs algorithm, Naïve Bayes classifier, an example: learning to classify text, Bayesian belief networks, the EM algorithm.

Computational learning theory: Introduction, probably learning an approximately correct hypothesis, sample complexity for finite hypothesis space, sample complexity for infinite hypothesis spaces, the mistake bound model of learning.

Instance - Based Learning: Introduction, *k*-nearest neighbor algorithm, locally weighted regression, radial basis functions, case-based reasoning, remarks on lazy and eager learning.

UNIT-IV

Genetic Algorithms: Motivation, Genetic algorithms, an illustrative example, hypothesis space search, genetic programming, models of evolution and learning, parallelizing genetic algorithms.

Learning Sets of Rules: Introduction, sequential covering algorithms, learning rule sets: summary, learning First-Order rules, learning sets of First-Order rules: FOIL, Induction as inverted deduction, inverting resolution.

Reinforcement Learning: Introduction, the learning task, *Q*–learning, non-deterministic, rewards and actions, temporal difference learning, generalizing from examples, relationship to dynamic programming.

UNIT-V

Analytical Learning - 1: Introduction, learning with perfect domain theories: PROLOG- EBG, remarks on explanation-based learning, explanation-based learning of search control knowledge.

Analytical Learning-2: Using prior knowledge to alter these arch objectives, using prior knowledge to augment search operators.

Combining Inductive and Analytical Learning – Motivation, inductive analytical approaches to learning, using prior knowledge to initialize the hypothesis.

TEXT BOOKS:

1. Tom M. Mitchell, "Machine Learning", McGraw-Hill, 1st Edition, 1997.
2. Kevin P. Murphy," Machine Learning: A Probabilistic Perspective", MIT Press, 1stEdition, 2012
3. Shai Shalev - Shwartz and Shai Ben-David, "Understanding Machine Learning: From Theory to Algorithms", Cambridge University Press, 1st Edition, 2014

REFERENCE BOOKS:

1. Stephen Marshland, "Machine Learning: An Algorithmic Perspective", Taylor &Francis, 2nd Edition, 2015.
2. Christopher M. Bishop, "Pattern Recognition and Machine Learning", CRC Press, 1st Edition, 2006.

BIG DATA ANALYTICS

Course Code: KG21CD603

L	T	P	C
3	0	0	3

B. Tech. III Year II - Semester

Prerequisites: A course on "Database Management Systems".

Course Objectives: The objectives of this course for the student are to:

1. Gain knowledge of Big data Analytics, principles and techniques.
2. Understand the frontiers of Big Data Technologies and Analytics.
3. Learn HADOOP framework and Map Reducing.
4. Understand HADOOP Architecture and Configuration.
5. Gain the knowledge of Data Analytics with R Machine Learning.

Course Outcomes: After completion of this course, the students will be able to

CO1: **Explain** the foundations, definitions, and challenges of Big Data and various Analytical tools.

CO2: **Apply** Big data technologies on parallel data source.

CO3: **Analyze** the programs using HADOOP, Map reduce and NO SQL.

CO4: **Justify** the importance of Big Data in Social Media and Mining applications.

CO5: **Analyze** Data Analytics for supervised and Unsupervised Learning using R Machine Learning.

UNIT-I

Introduction to Big Data: Big Data and its Importance – Four V's of Big Data Drivers for Big Data – Introduction to Big Data Analytics – Big Data Analytics applications.

UNIT-II

Big Data Technologies: Hadoop's Parallel World – Data discovery – Open source technology for Big Data Analytics – cloud and Big Data – Predictive Analytics – Mobile Business Intelligence and Big Data

UNIT-III

Introduction Hadoop: Big Data – Apache Hadoop & Hadoop Eco System – Moving Data in and out of Hadoop – Understanding inputs and outputs of Map Reduce – Data Serialization.

UNIT-IV

Hadoop Architecture: Hadoop: RDBMS Vs Hadoop, Hadoop Overview, Hadoop distributors, HDFS, HDFS Daemons, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, HDFS Architecture, Hadoop Configuration, Map Reduce Framework, Role of H Base in Big Data processing, HIVE, PIG.

UNIT-V

Data Analytics with R Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering, Social Media Analytics, Mobile Analytics, Big Data Analytics with Big R.

TEXT BOOKS:

1. Seema Acharya, Subhasini Chellappan, "Big Data Analytics", Wiley, 2015.

2. Michael Minelli, Michehe Chambers, Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", 1st Edition, Wiely CIO Series, 2013.

3. Tom White, "Hadoop: The Definitive Guide", 3rd Edition, O" Reilly Media, 2012.

4. Arvind Sathi, "Big Data Analytics: Disruptive Technologies for Changing the Game", 1st Edition, IBM Corporation, 2012.

REFERENCE BOOKS:

1. Jay Liebowitz, Auerbach Publications, "Big Data and Business Analytics", CRC press, 2013.

2. Tom Plunkett, Mark Hornick, "Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R Enterprise and Oracle R Connector for Hadoop", Mc Graw – Hill / Osborne Media, Oracle press, 2013.

3. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley, ISBN: 9788126551071, 2015.

4. Chris Eaton, Dirkderoosetal, "Understanding Big data", Mc Graw Hill, 2012.

5. Michael Berth old, David J. Hand, "Intelligent Data Analysis", Springer, 2007.

6. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", 1st Edition, Wiley and SAS Business Series, 2012.

DESIGN AND ANALYSIS OF ALGORITHMS

Course Code: KG21CS617

L	T	P	C
3	1	0	4

B. Tech. III Year II - Semester

Prerequisites:

Course Objectives: The objectives of this course for the student are to:

1. Introduces the notations for analysis of the performance of algorithms
2. Introduces the data structure disjoint sets.
3. Describes major algorithmic techniques (divide – and - conquer, backtracking, dynamic programming, greedy, branch and bound methods) and mention problems for which each technique is appropriate
4. Describes how to evaluate and compare different algorithms using worst-, average-, and best case analysis.
5. Explains the difference between tractable and intractable problems, andintroduces the problems that are P, NP and NP complete.

Course Outcomes: After completion of this course, the students will be able to

CO1: Analyze the complexity of algorithms by applying the knowledge ofasymptotic notations and recurrence methods.

CO2: Examine the performance of the Unions and Backtracking algorithms

CO3: Implement the given problem and identify appropriate algorithm design technique for problem solving.

CO4: Perceive and apply different algorithm design paradigms to findsolutions for computing problems.

CO5: Apply the knowledge of NP - hard and NP - Complete complexity classesto classify decision problems

UNIT-I

Introduction: Algorithm, Performance Analysis-Space complexity, Time complexity, Asymptotic Notations- Big oh notation, Omega notation, Theta notation and Little oh notation. Divide and conquer: General method, applications-Binary search, Quick sort, Merge sort, Strassen's matrix multiplication.

UNIT-II

Disjoint Sets: Disjoint set operations, union and find algorithms Backtracking: General method, applications, n-queen's problem, sum of subsets problem, graph coloring

UNIT-III

Dynamic Programming: General method, applications- Optimal binary search trees, 0/1 knapsack problem, All pairs shortest path problem, Traveling sales person problem, Reliability design.

UNIT-IV

Greedy method: General method, applications-Job sequencing with deadlines, knapsack problem, Minimum cost spanning trees, Single source shortest path problem.

UNIT-V

Branch and Bound: General method, applications - Travelling sales person problem, 0/1 knapsack problem - LC Branch and Bound solution, FIFO Branch and Bound solution. NP-Hard and NP-Complete problems: Basic

concepts, non deterministic algorithms, NP - Hard and NP-Complete classes, Cook's theorem

TEXT BOOKS:

1. Fundamentals of Computer Algorithms, Ellis Horowitz, Satraj Sahni and Rajasekharan, University Press.

REFERENCE BOOKS:

1. Design and Analysis of algorithms, Aho, Ullman and Hopcroft, Pearson education
2. Introduction to Algorithms, second edition, T. H. Cormen, C.E. Leiserson, R. L. Rivest, and C. Stein, PHI Pvt. Ltd./ Pearson Education
3. Algorithm Design: Foundations, Analysis and Internet Examples, M.T. Goodrich and R. Tamassia, John Wiley and sons

DATA VISUALIZATION TECHNIQUES

(Professional Elective - III)

Course Code: KG21CD615

L	T	P	C
3	0	0	3

B. Tech. III Year II - Semester

Prerequisites: A course on "Statistics with R"

Course Objectives: The objectives of this course for the student are to:

1. Understand various data visualization techniques.
2. Gain skills on both design and critique visualizations.
3. Understand role of visualization in data analysis.
4. Understand the components involved in visualization design.
5. Learn the data impacts in the type of visualization.

Course Outcomes: After completion of this course, the students will be able to

CO1: **Examine** and visualize the objects in different dimensions using visualization techniques.

CO2: **Design** and process the data for Virtualization.

CO3: **Apply** the visualization techniques in physical sciences, computer science, applied mathematics and medical science.

CO4: **Design** the virtualization techniques for research projects and applications.

CO5: **Solve** Problems by using visualization techniques.

UNIT-I

Introduction and Data Foundation: Basics - Relationship between Visualization and Other Fields - The Visualization Process - Pseudo code Conventions - The Scatter plot. Data Foundation - Types of Data - Structure within and between Records - Data Preprocessing - Data Sets.

UNIT-II

Foundations for Visualization: Visualization stages - Semiology of Graphical Symbols - The Eight Visual Variables - Historical Perspective - Taxonomies - Experimental Semiotics based on Perception Gibson's Affordance theory – A Model of Perceptual Processing.

UNIT-III

Visualization Techniques: Spatial Data: One-Dimensional Data - Two-Dimensional Data – Three Dimensional Data - Dynamic Data - Combining Techniques. Geospatial Data: Visualizing Spatial Data - Visualization of Point Data -Visualization of Line Data - Visualization of Area Data - Other Issues in Geospatial.

Data Visualization Multivariate Data: Point-Based Techniques - Line-Based Techniques - Region-Based Techniques - Combinations of Techniques – Trees Displaying Hierarchical Structures – Graphics and Networks-Displaying Arbitrary Graphs/Networks.

UNIT-IV

Interaction Concepts and Techniques: Text and Document Visualization: Introduction - Levels of Text Representations - The Vector Space Model - Single Document Visualizations -Document Collection Visualizations – Extended.

Text Visualizations Interaction Concepts: Interaction Operators - Interaction Operands and Spaces - A Unified Framework. Interaction

Techniques: Screen Space - Object-Space -Data Space -Attribute Space- Data Structure Space - Visualization Structure - Animating Transformations - Interaction Control.

UNIT-V

Research Directions in Virtualization: Steps in designing Visualizations – Problems in designing effective Visualizations- Issues of Data. Issues of Cognition, Perception, and Reasoning. Issues of System Design Evaluation, Hardware and Applications.

TEXT BOOKS:

1. Matthew Ward, Georges Grinstein and Daniel Keim, "Interactive Data Visualization Foundations, Techniques, Applications", 2010.
2. Colin Ware, "Information Visualization Perception for Design", 2nd Edition, Margon Kaufmann Publishers, 2004.

REFERENCE BOOKS:

1. Robert Spence "Information visualization – Design for interaction", Pearson Education, 2nd Edition, 2007.
2. Alexandru C. Telea, "Data Visualization: Principles and Practice," A. K. Peters LTD, 2008.