RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester

CD 501- Theory of Computation

COURSE OBJECTIVES:

This course will help students to learn several formal mathematical models of computation along with their relationships with formal languages and grammars. Students will also learn about solvable and unsolvable problems.

COURSE OUTCOMES:

After completing the course student should be able to:

1. Compare and analyze different theoretical computational models, languages and grammars.
2. Design and construct finite automata, pushdown automata and Turing machine for various problems.
3. Identify limitations of some computational models and possible methods of proving them.
4. Describe the concept of computable and non-computable problems.

Unit-I

Introduction of Automata Theory: Review of Sets, Mathematical formal proofs including proof by induction and by contradiction,Introduction to languages, grammars and automata: Alphabet, Representation of language and grammar, Types of Automata, Finite Automata as a language acceptor and translator, Moore machines and mealy machines, composite machine, Conversion from Mealy to Moore and vice versa.

Unit-II

Types of Finite Automata: Non Deterministic Finite Automata (NDFA), Deterministic finite automata machines, conversion of NDFA to DFA, minimization of automata machines, regular expression, applications of regular expressions, Arden's theorem. Meaning of union, intersection, concatenation and closure, 2 way DFA.

Unit-III

Grammars: Types of grammar, context sensitive grammar, and context free grammar, regular grammar. Derivation trees, ambiguity in grammar, simplification of context free grammar,conversion of grammar to automata machine and vice versa, Chomsky hierarchy of grammar, Chomsky normal form and Greibach normal form.

Unit-IV

Push down Automata: example of PDA, deterministic and non-deterministic PDAs, Context Free Grammar, Parsing, Ambiguity, Normal form of CFGs, CFG to NPDA, NPDA to CFGs CFG equivalent to PDA, Petri nets model.

Unit-V

Turing Machine: Turing Machine as acceptor, Recognizing a Language, Universal TMs, Linear Bounded Automata, Context Sensitive Languages, Recursive and Recursively Enumerable Languages, Unrestricted Grammars. Halting problem of Turing machine & the post correspondence problem, Concept of Solvability and Unsolvability, Church's Thesis, Complexity Theory – P and NP problems.

RECOMMENDED BOOKS

1. Hopcroft, Ullman, Motwani, "Introduction to Languages, Automata and Computation", 3rd Edition, Pearson Education, 2008.
2. John C. Martin, "Introduction to Languages and the Theory of Computation", Fourth Edition, Mc Graw Hill, 2010.
3. Peter Linz, "An Introduction to Formal Languages and Automata", Sixth Edition, Jones and Bartlett, 2016.
4. Lewis and Papadimitiriou, "Elements of Theory of Computation", Second Edition, Pearson Education, 2015.
5. K.L.P. Mishra and N. Chandrasekaran, "Theory of Computer Science: Automata, Languages and Computation", Third Edition, Prentice Hall, 2006.
6. Cohen John, "Introduction to Computer Theory", Second Edition, Wiley and Sons, 2007.
7. Theory of Computation, Wood, Harper & Row.

LIST OF EXPERIMENTS

1. Design a Program for creating machine that accepts three consecutive one.
2. Design a Program for creating machine that accepts the string always ending with 101.
3. Design a Program for Mode 3 Machine
4. Design a program for accepting decimal number divisible by 2.
5. Design a program for creating a machine which accepts string having equal no. of 1's and 0's.
6. Design a program for creating a machine which count number of 1's and 0's in a given string.
7. Design a Program to find 2's complement of a given binary number.
8. Design a Program which will increment the given binary number by 1.
9. Design a Program to convert NDFA to DFA.
10. Design a Program to create PDA machine that accept the well-formed parenthesis.
11. Design a PDA to accept WCWR where w is any string and WR is reverse of that string and C is a Special symbol.
12. Design a Turing machine that's accepts the following language $a^nb^nc^n$ where $n>0$.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester
CD502-Machine Learning

COURSE OBJECTIVES:

The objective of this course is to impart necessary knowledge of different machine learning techniques and develop programming skills required to build machine learning based applications.

COURSE OUTCOMES:

After completing the course student should be able to:

1. Describe in-depth about theories, methods, and algorithms in machine learning.
2. Find and analyze the optimal hyper parameters of the machine learning algorithms.
3. Examine the nature of a problem at hand and determine whether machine learning can solve it efficiently.
4. Solve and implement real world problems using machine learning.

COURSE CONTENTS:

UNIT-I

Introduction to machine learning, Machine learning life cycle, Types of Machine Learning System (supervised and unsupervised learning, Batch and online learning, Instance-Based and Model based Learning), scope and limitations, Challenges of Machine learning, data visualization, hypothesis function and testing, data pre-processing, data augmentation, normalizing data sets, , Bias-Variance tradeoff, Relation between AI (Artificial Intelligence), ML (Machine Learning), DL (Deep Learning) and DS (Data Science).

UNIT-II

Clustering in Machine Learning: Types of Clustering Method: Partitioning Clustering, Distribution Model-Based Clustering, Hierarchical Clustering, Fuzzy Clustering. Birch Algorithm, CURE Algorithm. Gaussian Mixture Models and Expectation Maximization. Parameters estimations – MLE, MAP. Applications of Clustering.

UNIT-III

Classification algorithm: - Logistic Regression, Decision Tree Classification, Neural Network, K-Nearest Neighbors (K-NN), Support Vector Machine, Naive Bayes (Gaussian, Multinomial, Bernoulli). Performance Measures: Confusion Matrix, Classification Accuracy, Classification Report: Precisions, Recall, F1 score and Support.

UNIT-IV

Ensemble Learning and Random Forest**:** Introduction to Ensemble Learning, Basic Ensemble Techniques (Max Voting, Averaging, Weighted Average), Voting Classifiers, Bagging and Pasting, Out-of-Bag Evaluation, Random Patches and Random Subspaces, Random Forests (Extra-Trees, Feature Importance), Boosting (AdaBoost, Gradient Boosting), Stacking.

UNIT-V

Dimensionality Reduction:The Curse of Dimensionality, Main Approaches for Dimensionality Reduction (Projection, Manifold Learning) PCA: Preserving the Variance, Principal Components, Projecting Down to d Dimensions, Explained Variance Ratio, Choosing the Right Number of Dimensions, PCA for Compression, Randomized PCA, Incremental PCA. Kernel PCA: Selecting a Kernel and Tuning Hyper parameters. Learning Theory: PAC and VC model.

REFERENCE BOOKS:

1. Tom M. Mitchell, “Machine Learning”, McGraw Hill Education, First edition, 2017.
2. Aurelien Geon, “Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems”, Shroff/O'Reilly; First edition (2017).
3. Andreas Muller, "Introduction to Machine Learning with Python: A Guide for Data Scientists", Shroff/O'Reilly; First edition (2016).
4. Leonard Kaufman and P. J. Rousseau. Finding groups in data: An introduction to cluster analysis, Wiley, 2005
5. NelloCristianini and John Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.

PRACTICAL:

Different problems to be framed to enable students to understand the concept learnt and get hands-on on various tools and software related to the subject. Such assignments are to be framed for ten to twelve lab sessions.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based on AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester

CD503 (A) Data Mining & Warehousing

COURSE OBJECTIVES:

Student should understand the value of Historical data and data mining in solving real-world problems. Student should become affluent with the basic Supervised and unsupervised learning algorithms commonly used in data mining.
Student develops the skill in using data mining for solving real-world problems.

Unit 1: Data Warehousing: Introduction, Delivery Process, Data warehouse Architecture, Data Preprocessing: Data cleaning, Data Integration and transformation, Data reduction. Data warehouse Design: Dataware house schema, Partitioning strategy Data warehouse Implementation, Data Marts, Meta Data, Example of a Multidimensional Data model, Introduction to Pattern Warehousing.

Unit 2: OLAP Systems: Basic concepts, OLAP queries, Types of OLAP servers, OLAP operations etc. Data Warehouse Hardware and Operational Design: Security, Backup And Recovery,

Unit 3: Introduction to Data & Data Mining: Data Types, Quality of data, Data Preprocessing, Similarity measures, Summary statistics, Data distributions, Basic data mining tasks, Data Mining V/s knowledge discovery in databases. Issues in Data mining, Introduction to Fuzzy sets and fuzzy logic.

Unit 4: Supervised Learning (Classification): Statistical-based algorithms, Distance-based algorithms, Decision tree-based algorithms, Neural network-based algorithms, Rule-based algorithms, Probabilistic Classifiers

Unit 5: Clustering & Association Rule mining: Hierarchical algorithms, Partitional algorithms, Clustering large databases – BIRCH, DBSCAN, CURE algorithms. Association rules : Parallel and distributed algorithms such as Apriori and FP growth algorithms.

Books Recommended:

Text Books:

1. Pang – ningTan , Steinbach & Kumar, “Introduction to Data Mining”, Pearson Edu, 2019.
2. Jaiwei Han, MichelineKamber, “Data Mining : Concepts and Techniques”, Morgan Kaufmann Publishers.

Reference Books:

1. Margaret H. Dunham, “Data Mining : Introductory and Advanced topics”, Pearson Edu., 2009.
2. Anahory& Murray, “Data Warehousing in the Real World”, Pearson Edu., 2009.

COURSE OUTCOMES

After completion of this course, the students would be able to:

- CO1. Understand the need of designing Enterprise data warehouses and will be enabled to approach business problems analytically by identifying opportunities to derive business.
- CO2. Compare and contrast various methods for storing & retrieving data from different data sources/repository.
- CO3. Ascertain the application of data mining in various areas and Preprocess the given data and visualize it for a given application or data exploration/mining task
- CO4. Apply supervised learning methods to given data sets such as classification and its various types.
- CO5. Apply Unsupervised learning methods to given data sets such as clustering and its various types.
- CO6. Apply Association rule Mining to various domains.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based on AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester

CD503 (B) PATTERN RECOGNITION

Unit-I

Introduction – Definitions, datasets for Pattern, Application Areas and Examples of pattern recognition, Design principles of pattern recognition system, Classification and clustering, supervised Learning, unsupervised learning and adaptation, Pattern recognition approaches, Decision Boundaries, Decision region , Metric spaces, distances.

Unit -II

Classification: introduction, application of classification, types of classification, decision tree, naïve bayes, logistic regression , support vector machine, random forest, K Nearest Neighbour Classifier and variants, Efficient algorithms for nearest neighbour classification, Different Approaches to Prototype Selection, Combination of Classifiers, Training set, test set, standardization and normalization.

Unit – III

Different Paradigms of Pattern Recognition, Representations of Patterns and Classes, Unsupervised Learning & Clustering: Criterion functions for clustering, Clustering Techniques: Iterative square -error partitional clustering – K means, hierarchical clustering, Cluster validation.

Unit -IV

Introduction of feature extraction and feature selection, types of feature extraction , Problem statement and Uses, Algorithms - Branch and bound algorithm, sequential forward / backward selection algorithms, (l,r) algorithm.

Unit -V

Recent advances in Pattern Recognition, Structural PR, SVMs, FCM, Soft computing and Neuro-fuzzy techniques, and real-life examples, Histograms rules, Density Estimation, Nearest Neighbor Rule, Fuzzy classification.

REFERENCE BOOKS:

1. Richard O. Duda, Peter E. Hart and David G. Stork, "Pattern Classification", 2nd Edition, John Wiley, 2006.
2. C. M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2009.
3. S. Theodoridis and K. Koutroumbas, "Pattern Recognition", 4th Edition, academic Press, 2009.
4. Robert Schalkoff, "Pattern Recognition: statistical, structural and neural approaches", JohnWiley&sons ,Inc, 2007.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based on AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester

CD503 (C) Introduction to Toolkits for Data Science

Course Content:

Unit 1: Python for Data Science: Review of Numpy, Pandas and Scikit-learn.Supervised Learning Techniques packages/toolkit for regression and classification: - Decision Trees, Naive Bayes, Classification, Support vector machines, Random Forest, Neural network, Ensemble Methods, Ordinary Least Squares Regression, Logistic Regression, etc. Unsupervised Learning, Clustering: k-means, adaptive hierarchical clustering, Gaussian mixture, Optimization Using Evolutionary Techniques etc.

Unit 2: R for Data Science: Basic of R and RStudio. R data structures: vectors, factors, lists, arrays, matrices, and data frames. Working with data: Import data into R and visualize data. Data Analytics Software: Weka, Orange, Rapidminer, Minitab, PowerBI, GitHub, Google Colab.

Unit 3. Introduction to Deep Learning: Basics of TensorFlow and keras,Basics of PyTorch, perform style transfer of one image to another, Perform text generation, and sentiment analysis with PyTorch. Neural networks that recognize objects, improve the accuracy of object recognition using CNN, use pre-trained models to build state-of-the-art classifiers, Saving and Loading models, Time series forecasting with RNNs, and LSTMs,

Unit 4: Introduction to Time Series Analysis: Time series regression and exploratory data analysis toolkits: ARMA/ARIMA models, model identification/estimation/linear operators, Fourier analysis, spectral estimation, and state-space models.

Unit 5: Cloud Computing for Data Science: Implementation of Machine Learning and Deep learning through AWS/Azure platform.

Version controlling tools for data science projects. Case studies of data science projects.

Books Recommended:

- Brockwell& Davis (2016) Introduction to Time Series and Forecasting, 3rd edition, Springer
- Cryer& Chan (2008) Time-Series Analysis with Applications in R, Springer
- Prado & West (2010) Time Series: Modeling, Computation, and Inference Chapman & Hall
- Petris, Petrone, Campagnoli (2009) Dynamic Linear Models with R, Springer
- Ruppert& Matteson (2016) Statistics and Data Analysis for Financial Engineering with R examples, 2nd Edition, Springer
- R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, 1st Edition, O'reilly publication.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based on AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester

CD504 (A) Computer Graphics & Visualization

Unit-I Introduction to Raster Scan displays, Pixels, Frame buffer, Vector & Character generation, Random Scan systems, Display devices, Scan Conversion techniques, Line Drawing algorithms: simple DDA, Bresenham's Algorithm, Circle Drawing Algorithms: Midpoint Circle drawing and Bresenham's Algorithm, Polygon fill algorithm: Boundary-fill and Flood-fill algorithms.

Unit-II 2-D Transformation: Translation, Rotation, Scaling, Shearing, Reflection. Inverse Transformation, Homogeneous coordinate system, Matrices Transformation, Composite Transformation. Windowing & Clipping: World Coordinate System, Screen Coordinate System, Viewing Transformation, Line Clipping & Polygon Clipping Algorithms

Unit-III 3-D Transformations: Translation, Rotation and Scaling. Parallel & Perspective Projection: Types of Parallel & Perspective Projection, Hidden Surface elimination: Depth comparison, Back face detection algorithm, Painter's Algorithm, Z-Buffer Algorithm. Curve generation, Bezier and B-spline methods. Basic Illumination Model: Diffuse reflection, Specular reflection, Phong Shading, Gouraud shading, Ray Tracing, Color models like RGB, YIQ, CMY, HSV.

Unit-IV Visualization: Visualization of 2D/3D scalar fields: color mapping, ISO surfaces. Direct volume data rendering: ray-casting, transfer functions, segmentation. Visualization of Vector fields and flow data, Time-varying data, High-dimensional data: dimension reduction, parallel coordinates, Non-spatial data: multi-variate, tree/graph structured, text Perceptual and cognitive foundations, Evaluation of visualization methods, Applications of visualization, Basic Animation Techniques like traditional, key framing

Unit –V Multimedia :Basic of multimedia, application of Multimedia, Text-Types, Unicode Standard ,text Compression, Text file formats, Audio Components, Digital Audio, Digital Audio processing, Sound cards, Audio file formats ,Audio Processing software ,Video-Video color spaces, Digital Video, Digital Video processing, Video file formats. Animation: Uses of Animation, Principles of Animation, Computer based animation, 3D Animation, Animation file formats, Animation software,Special Effects in animation, Storyboarding for Animation, Compression: Lossless/Lossy Compression techniques, Image, Audio & Video Compression, MPEG Standards ,Multimedia Architecture, Multimedia databases.

Recommended Text:

1. Donald Hearn and M.P. Becker "Computer Graphics" Pearson Pub.
2. Foley, Van Dam, Feiner, Hughes, "Computer Graphics: Principles and Practice" Addison- Wesley

3. Rogers, "Procedural Elements of Computer Graphics", Tata McGraw Hill
4. Parekh “Principles of Multimedia” Tata McGraw Hill
5. Maurya, “Computer Graphics with Virtual Reality System “ , Wiley India
6. Pakhira,”Computer Graphics ,Multimedia &Animation”,PHI learning
7. Andleigh, Thakral , “Multimedia System Design “ PHI Learning
8. Khalid Sayood , “Introduction to Data Compression”, Morgan Kaufmann

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based on AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester

CD504 (B) Data Compression

COURSE DESCRIPTION:

The course covers the theory of quantization and basic concepts in source coding and applications of the theory and concepts to systems that convert analog or high-rate digital signals into lowrate digital representations with or without loss of fidelity. The concept of source coding is extended to general descriptions of a statistical information source where various data modeling techniques find useful applications.

COURSE OBJECTIVES

The objective of this course is to

1. Gain a fundamental understanding of data compression methods for text, images, and video, and related issues in the storage, access, and use of large data sets
2. Select, giving reasons that are sensitive to the specific application and particular circumstance, most appropriate compression techniques for text, audio, image and video information
3. Illustrate the concept of various algorithms for compressing text, audio, image and video information.

COURSE OUTCOMES

On completion of this course, the students will be able to:

1. program, analyze Hoffman coding: Lossless image compression, Text compression, Audio Compression
2. program and analyze various Image compression and dictionary based techniques like static Dictionary, Diagram Coding, Adaptive Dictionary
3. understand the statistical basis and performance metrics for lossless compression
4. understand the conceptual basis for commonly used lossless compression techniques, and understand how to use and evaluate several readily available implementations of those techniques
5. understand the structural basis for and performance metrics for commonly used lossy compression techniques and conceptual basis for commonly used lossy compression techniques.

COURSE CONTENT

Unit I: Compression Techniques

Loss less compression, Lossy Compression, Measures of performance, Modeling and coding, Mathematical Preliminaries for Lossless compression: A brief introduction to information theory, Models: Physical models, Probability models, Markov models, com-posite source model, Coding: uniquely decodable codes, Prefix codes.

Unit II: The Huffman coding algorithm

Minimum variance Huffman codes, Adaptive Huffman coding: Update procedure, Encoding procedure, Decoding procedure. Golomb codes, Rice codes, Tunstall codes, Applications of Hoffman coding: Lossless image compression, Text compression, Audio Compression.

Unit III: Coding

Coding a sequence, Generating a binary code, Comparison of Binary and Huffman coding, Applications: Bi-level image compression- The JBIG standard, JBIG2, Image compression. Dictionary Techniques: Introduction, Static Dictionary: Diagram Coding, Adaptive Dictionary. The LZ77 Approach, The LZ78 Approach, Applications: File Compression-UNIX compress, ImageCompression: The Graphics Interchange Format (GIF),

Compression over Modems: V.42 bits, Predictive Coding:

Prediction with Partial match (ppm): The basic algorithm, The ESCAPE SYMBOL, length of context, The Exclusion Principle, The Burrows-Wheeler Transform: Move to- front coding, CALIC, JPEGLS, Multi-resolution Approaches, Facsimile Encoding, Dynamic Markoy Compression.

Unit IV: Scalar Quantization

Distortion criteria, Models, Scalar Quantization: The Quantization problem, Uniform Quantizer, Adaptive Quantization, Non uniform Quantization.

Unit V: Vector Quantization

Advantages of Vector Quantization over Scalar Quantization, TheLinde-Buzo-Gray Algorithm.

TEXT BOOKS

1. The Data Compression Book – Mark Nelson.
2. Data Compression: The Complete Reference – David Salomon.

REFERENCE BOOKS

1. Introduction to Data Compression – Khalid Sayood, MorganKaufmann Publishers.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based on AICTE Flexible Curricula

CSE-Data Science/Data Science, V semester

CD504 (C) Computer Organization & Architecture

Unit I:

Review of Digital Logic Circuits, Digital Logic Components and Data representation.Computer Arithmetic**:** Integer and Floating Point Arithmetic operations.
Computer Organization v/s Architecture,Milestonesin Computer Architecture, Basic Structure of Computer System, Componentsof Computer System-CPU;Memory;System Bus- Bus width, Bus Operations;I/O subsystem. CPU Organization: General Register Organization-Memory Register, Instruction Register; Control Word, Stack Organization; ALU, Control Unit.

Unit II:

(A) Machine Language Level/Instruction Set Architecture (ISA) level: Instruction Set-Machine Instruction Characteristics, Types of operands, Types of operations**;** Instruction Types, Instruction Formats, Addressing Modes; Registers, Program Counter; Instruction Execution Cycle;Interrupts and Traps, Sources of interrupts, Interrupt identification and priorities, Interrupt servicing.Case Study of 8086 Microprocessor.
(B) Control Unit: Hardwired Control Unit;Micro-programmed Control Unit-Micro Instructions, Micro Instruction Formats, Micro Instruction Control, Micro program sequencer, Execution of Micro Instructions.

Unit III:

Memory Organization: Memory Hierarchy, Main memory-RAM, ROM; Memory Technologies; Memory Addresses, Memory Address Map; Flash Memory; Associative Memory, Cache Memory: Cache Structure and Design, Mapping Schemes, Replacement Algorithms, Improving Cache Performance; Concept of L1, L2, L3 Cache.
Secondary Memory –Magnetic Tape, Magnetic Disk, Optical Disks, Solid State Disk.

Unit IV:

I/O Organization: Data Transfer- Serial, Parallel, Synchronous, Asynchronous Modes of Data Transfer, I/O Techniques- Programmed I/O, Interrupt driven I/O,Direct Memory Access(DMA); External Interconnection Standards **(**I/O Interfaces): PCI Bus,PCI Express, SCSI Bus, USB; I/O Channels and I/O Processors; I/O Instructions**.**

Unit V:

Parallel Architectures:On-chip parallelism, Thread level parallelism, Instruction level parallelism; Multicore Processor Architecture;Processor level parallelism; Overview of Pipelining, Vector Processing and Array Processing. RISC vs CISC Architectures. Introduction to ARM processor and its architecture. Introduction to Assembly Language Programming.

Books Recommended:

1. William Stallings, “Computer Organization and architecture”, Pearson.
2. Tannenbaum and Austin, “Structured Computer Organization”, PHI.
3. V. Carl Hamacher, “Computer Organization”, McGraw Hill.
4. John P. Hayes, “Computer Architecture and Organization”, TMH.
5. Morris Mano, “Computer System Architecture”, PHI.
6. David A. Patterson and John L. Hennessy, “Computer Organization and Design: The Hardware/Software Interface”, Morgan Kauffman.
7. M. Usha, T.S. Shrikant: “Computer System Architecture and Organization”, Willey India.
8. Chaudhuri, P.Pal: “Computer Organization and Design”,PHI
9. Sarangi: “Computer Organization and Architecture”, McGraw Hill.