Syllabus of

UNDERGRADUATE DEGREE COURSE

B.Tech. V Semester

Computer Science & Engineering

(Data Science)

Rajasthan Technical University, Kota

Effective from session: 2021-22

5CDS-01: Data Mining-Concepts and Techniques

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

Course Objectives:

1. To introduce the fundamental processes data warehousing and major issues in data mining
2. To impart the knowledge on various data mining concepts and techniques that can be applied to text mining, web mining etc.
3. To develop the knowledge for application of data mining and social impacts of data mining.

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

1. Interpret the contribution of data warehousing and data mining to the decision-support systems.
2. Prepare the data needed for data mining using preprocessing techniques.
3. Extract useful information from the labeled data using various classifiers.
4. Compile unlabeled data into clusters applying various clustering algorithms.
5. Discover interesting patterns from large amounts of data using Association Rule Mining
6. Demonstrate capacity to perform a self-directed piece of practical work that requires the application of data mining techniques.

Detailed Syllabus: (per session plan)

UNIT	Contents
1	Introduction to Data Mining: Introduction to data mining-Data mining functionalities-Steps in data mining process- Classification of data mining systems, Major issues in data mining. Data Wrangling and Preprocessing: Data Preprocessing: An overview-Data cleaning-Data transformation and Data discretization
2	Predictive Modeling: General approach to classification-Decision tree induction-Bayes classification methods- advanced classification methods: Bayesian belief networks- Classification by Backpropagation- Support Vector Machines-Lazy learners
3	Descriptive Modeling: Types of data in cluster analysis-Partitioning methods-Hierarchical methods-Advanced cluster analysis: Probabilistic model-based clustering- Clustering high- dimensional data-Outlier analysis
4	Discovering Patterns and Rules: Frequent Pattern Mining: Basic Concepts and a Road Map - Efficient and scalable frequent item set mining methods: Apriori algorithm, FP-Growth algorithm- Mining frequent itemsets using vertical data format- Mining closed and max patterns- Advanced Pattern Mining: Pattern Mining in Multilevel, Multidimensional Space
5	Data Mining Trends and Research Frontiers: Other methodologies of data mining: Web mining- Temporal mining-Spatial mining-Statistical data mining-Visual and audio data mining- Data mining applications- Data mining and society: Ubiquitous and invisible data mining- Privacy, Security, and Social Impacts of data mining

TEXT BOOKS:

1. Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, third edition ,2013
2. Pang-Ning Tan,Michael Steinbach, Anuj Karpatne, Vipin Kumar, Introduction to Data Mining, second edition, Pearson, 2019

REFERENCE BOOKS:

1. Ian.H.Witten, Eibe Frank and Mark.A.Hall, Data Mining:Practical Machine Learning Tools and Techniques,third edition , 2017
2. Alex Berson and Stephen J. Smith, Data Warehousing, Data Mining & OLAP, Tata McGraw Hill Edition, Tenth Reprint, 2008.
3. Hand, D., Mannila, H. and Smyth, P. Principles of Data Mining, MIT Press: Massachusets. third edition, Pearson, 2013

5CDS-02: Compiler Design

Credit: 3 **Max. Marks: 100(IA:30, ETE:70)**

3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Introduction:** Objective, scope and outcome of the course. Compiler, Translator, Interpreter definition, Phase of compiler, Bootstrapping, Review of Finite automata lexical analyzer, Input, Recognition of tokens, Idea about LEX: A lexical analyzer generator, Error handling.	06
3	**Review of CFG Ambiguity of grammars:** Introduction to parsing. Top down parsing, LL grammars & passers error handling of LL parser, Recursive descent parsing predictive parsers, Bottom up parsing, Shift reduce parsing, LR parsers, Construction of SLR, Conical LR & LALR parsing tables, parsing with ambiguous grammar. Operator precedence parsing, Introduction of automatic parser generator: YACC error handling in LR parsers.	10
4	**Syntax directed definitions;** Construction of syntax trees, S-Attributed Definition, L-attributed definitions, Top down translation. Intermediate code forms using postfix notation, DAG, Three address code, TAC for various control structures, Representing TAC using triples and quadruples, Boolean expression and control structures.	10
5	**Storage organization;** Storage allocation, Strategies, Activation records, Accessing local and non-local names in a block structured language, Parameters passing, Symbol table organization, Data structures used in symbol tables.	08
6	**Definition of basic block control flow graphs;** DAG representation of basic block, Advantages of DAG, Sources of optimization, Loop optimization, Idea about global data flow analysis, Loop invariant computation, Peephole optimization, Issues in design of code generator, A simple code generator, Code generation from DAG.	07
	Total	42

5CDS-03: Operating System

Credit: 3 **Max. Marks: 100(IA:30, ETE:70)**
3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Introduction and History of Operating systems:** Structure and operations; processes and files Processor management: inter process communication, mutual exclusion, semaphores, wait and signal procedures, process scheduling and algorithms, critical sections, threads, multithreading	04
3	**Memory management:** contiguous memory allocation, virtual memory, paging, page table structure, demand paging, page replacement policies, thrashing, segmentation, case study	05
4	**Deadlock:** Shared resources, resource allocation and scheduling, resource graph models, deadlock detection, deadlock avoidance, deadlock prevention algorithms **Device management:** devices and their characteristics, device drivers, device handling, disk scheduling algorithms and policies	15
5	**File management:** file concept, types and structures, directory structure, cases studies, access methods and matrices, file security, user authentication	07
6	**UNIX and Linux operating systems as case studies;** Time OS and case studies of Mobile OS	08
	Total	40

5CDS-04: Data visualization- R Programming/ Power BI

Credit: 3 **Max. Marks: 100(IA:30, ETE:70)**
3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Introduction to Data Science and Data Visualization:** Data Science Introduction: Concepts, lifecycle, applications. - Role of Data Visualization in Analysis and Decision Making. - Basics of R Programming: Variables, data types, operators. - Data Visualization Fundamentals: Principles, visualization types.	07
3	**Data Preprocessing and EDA with R:**Data Collection and Sources: Structured, unstructured, web scraping. - Data Cleaning: Handling missing data, outliers. - Data Transformation Techniques: Normalization, standardization, encoding. - Exploratory Data Analysis (EDA): Univariate, bivariate, multivariate analysis. - Advanced EDA Plotting: ggplot2 for customized visualizations, faceting, distributions.	10
4	**Advanced Data Analysis and Visualization with R:** - Statistical Analysis: Descriptive stats, hypothesis testing. - Data Visualization Libraries in R: ggplot2. - Machine Learning Concepts: Introduction to ML, basic models in R. - R Shiny: Building interactive web applications.	08
5	**Power BI for Data Visualization and Dashboard Creation:** Introduction to Power BI: Interface, data connection, roles. - Creating Basic Visualizations: Bar charts, line charts, scatter plots. - Building Interactive Dashboards: Design principles, combining visualizations. - Effective Data Storytelling using Power BI.	08
6	**Advanced Data Visualization and Integration:** Advanced Visualization Techniques in R. - Integrating R with Power BI: Using R scripts and calculations. - Data Visualization Ethics and Best Practices. - Capstone Project: Applying skills using R and Power BI.	08
	Total	41

Textbooks

- R for Data Science by Hadley Wickham and Garrett Grolemund
- Hands-On Data Visualization with R by Claus O. Wilke
- Power BI for Data Science by Ryan Sleeper
- Data Visualization with Power BI by Daniel Murray
- Data Storytelling with Power BI by Anupam Jain

Reference Books

- The Visual Display of Quantitative Information by Edward Tufte
- Data Visualization: A Practical Introduction by Kieran Healy
- The Functional Art of Data Visualization by Alberto Cairo
- Storytelling with Data by Cole Nussbaumer Knaflic
- Information Visualization: Perception for Design by Colin Ware

5CDS-05: Analysis of Algorithms

Credit: 3 **Max. Marks: 100(IA:30, ETE:70)**
3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Background**: Review of Algorithm, Complexity Order Notations: definitions and calculating complexity. **Divide And Conquer Method**: Binary Search, Merge Sort, Quick sort and Strassen's matrix multiplication algorithms.	06
3	**Greedy Method**: Knapsack Problem, Job Sequencing, Optimal Merge Patterns and Minimal Spanning Trees. **Dynamic Programming**: Matrix Chain Multiplication. Longest Common Subsequence and 0/1 Knapsack Problem.	10
4	**Branch And Bound**: Traveling Salesman Problem and Lower Bound Theory. Backtracking Algorithms and queens problem. **Pattern Matching Algorithms**: Naïve and Rabin Karp string matching algorithms, KMP Matcher and Boyer Moore Algorithms.	08
5	**Assignment Problems:** Formulation of Assignment and Quadratic Assignment Problem. **Randomized Algorithms-** Las Vegas algorithms, Monte Carlo algorithms, randomized algorithm for Min-Cut, randomized algorithm for 2- SAT. Problem definition of Multi commodity flow, Flow shop scheduling and Network capacity assignment problems.	08
6	**Problem Classes Np, Np-Hard And Np-Complete:** Definitions of P, NP-Hard and NP-Complete Problems. Decision Problems. Cook's Theorem. Proving NP- Complete Problems - Satisfiability problem and Vertex Cover Problem. Approximation Algorithms for Vertex Cover and Set Cover Problem.	08
	Total	41

5CDS-11: Fundamentals of Block chain

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

Course Objectives:

1. The students should be able to understand a broad overview of the essential concepts of block chain technology.
2. To familiarize students with Bitcoin protocol followed by the Ethereum protocol – to lay the foundation necessary for developing applications and programming.
3. Students should be able to learn about different types of block chain and consensus algorithms.

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

1. To explain the basic notion of distributed systems.
2. To use the working of an immutable distributed ledger and trust model that defines block chain.
3. To illustrate the essential components of a block chain platform.

Detailed Syllabus: (per session plan)

UNIT	Contents
1	Basics: The Double-Spend Problem, Byzantine Generals' Computing Problems, Public-Key Cryptography, Hashing, Distributed Systems, Distributed Consensus.
2	Technology Stack: Block chain, Protocol, Currency. Bitcoin Block chain: Structure, Operations, Features, Consensus Model, Incentive Model
3	Ethereum Block chain: Smart Contracts, Ethereum Structure, Operations, Consensus Model, Incentive Model.
4	Tiers of Block chain Technology: Block chain 1.0, Block chain 2.0, Block chain 3.0, Types of Block chain: Public Block chain, Private Block chain, Semi-Private Block chain, Sidechains.
5	Types of Consensus Algorithms: Proof of Stake, Proof of Work, Delegated Proof of Stake, Proof Elapsed Time, Deposite-Based Consensus, Proof of Importance, Federated Consensus or Federated Byzantine Consensus, Practical Byzantine Fault Tolerance. Block chain Use Case: Supply Chain Management.

TEXT BOOKS:

1. Kirankalyan Kulkarni, Essentials of Bitcoin and Blockchain, Packt Publishing.
2. Anshul Kaushik, Block Chain & Crypto Currencies, Khanna Publishing House.
3. Tiana Laurence, Blockchain for Dummies, 2nd Edition 2019, John Wiley & Sons.
4. Mastering Blockchain: Deeper insights into decentralization, cryptography, Bitcoin, and popular Blockchain frameworks by Imran Bashir, Packt Publishing (2017).

REFERENCE BOOKS:

1. Blockchain: Blueprint for a New Economy by Melanie Swan, Shroff Publisher O'Reilly Publisher Media; 1st edition (2015).
2. Mastering Bitcoin: Programming the Open Blockchain by Andreas Antonopoulos.

Syllabus

5CDS-12: Probability and Statistics for Data Science

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**

2L+0T+0P **End Term Exam: 3 Hours**

Objectives:

1. To provide advanced statistical background for analyzing data and drawing inferences from that analysis.
2. Predictive Analytics using linear and generalized linear models.

Outcomes:

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

- Students will be able to learn advanced statistical techniques and apply them to the analysis of real data sets from different fields.

Detailed Syllabus: (per session plan)

UNIT	Contents
1	Descriptive Statistics: a) Measures of Central Tendencies – Grouped and Ungrouped Data; Mean, Sample Mean– Weighted mean; Median, Quartiles, b) Deciles and Percentiles, Box plot, Mode Measures of Variability– Dispersion, Range, Standard deviation, Population v/s sample variance and standard deviation, Skewness, Kurtosis.
2	Introduction to Probability and Sampling distribution: a) Methods of Assigning probabilities, Probability Space, conditions of probability model, Events, simple and compound, Laws of probability, Probability density function, Cumulative distribution function, Expected values of Mean and Variance. Marginal, union, joint and conditional probabilities, Bayes' Theorem b) Random variables, discrete and continuous distributions, Expectation, moments of a distribution, Binomial, Poisson, uniform, and normal distributions, Normal approximation to the binomial distribution, Distributions of several random variables, moments of joint distributions, independence, covariance, correlation coefficient, Central Limit Theorem.
3	Hypothesis Testing: a) Large Sample estimation of the population parameters and Hypothesis testing: Basics of Estimating the populations mean and difference; estimating the proportion and difference; large sample test for population mean, difference; large sample test for proportion, difference. b) Estimation of a population variance: Sampling distribution of variance, estimation. c) Inferences from small sample: Student's *t* distribution; Small sample *t* test for following – A population mean, A difference between two means, Confidence interval.
4	Regression Model: a) least squares and linear regression: Introduction; Notation; Ordinary least squares; Regression to the mean; Linear regression; Residuals; Regression inference b) Multivariable regression: Multivariate regression; Multivariate examples; Adjustment; Residual variation and diagnostics; Multiple variables, Interaction Terms, Non-linear Transformations of the Predictors, Qualitative Predictors.

5	Generalized linear models: Logistic Regression, Binary outcomes, Count outcomes, Multiple Logistic Regression ANOVA/MANOVA: Chi-Square and Analysis of Variance, Multivariate analysis of variance Extension of regression analysis: Ridge Regression, The Lasso

Text Books:

1. An Introduction to Statistical Learning with application in R . Hastie T, Robert T. (2014). Springer Science Business Media: New York

Reference Books:

1. Statistics for Management, Seventh Edition, by Richard I. Levin, David S. Rubin, Pearson
2. An Introduction to Categorical Data Analysis. Agresti, A. (2012). John Wiley & sons
3. The Element of Statistical Learning, Data mining, Inference and Prediction. Hastie, T, Tibshirani, R, & Friedman, J. (2011). New York: Springer Series in Statistics.
4. Hair, Black, Babin, Anderson and Tatham (2009). Multivariate Data Analysis, Pearson

B.Tech.: Computer Science & Data Science

5CDS-13: Programming for Data Science

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**

2L+0T+0P **End Term Exam: 3 Hours**

Course Objectives:		
1. To provide necessary knowledge on data manipulation and to perform analysis on the practical problems using statistical and machine learning approach 2. To generate report and visualize the results in graphical form using programming tool		
Expected Course Outcome:		
1. Ability to gain basic knowledge on data science 2. Convert the real time data into suitable form for analysis 3. Gain the insights from the data through statistical inferences 4. Develop suitable models using machine learning techniques and to analyze its performance 5. Identify the requirement and visualize the results 6. Analyze on the performance of the model and the quality of the results		
Unit:1	**INTRODUCTION**	**4 hours**
Data Science: Introduction to Data Science – Digital Universe – Sources of Data – Information Commons – Data Science Project Life Cycle: OSEMN Framework		
Unit:2	**DATA PREPROCESSING & CONCEPT LEARNING**	**6 hours**
Introduction to Data Preprocessing – Reading, Selecting, Filtering Data – Filtering Missing Values – Manipulating, Sorting, Grouping, Rearranging, Ranking Data Formulation of Hypothesis – Probabilistic Approximately Correct Learning - VC Dimension – Hypothesis elimination – Candidate Elimination Algorithm		
Unit:3	**ESSENTIALS OF R**	**8 hours**
R Basics - data types and objects - control structures – data frame -Feature Engineering - scaling, Label Encoding and One Hot Encoding, Reduction		
Unit:4	**MODEL FIT USING R**	**8 hours**
Regression Models- Linear and Logistic Model, Classification Models – Decision Tree, Naïve Bayes, SVM and Random Forest, Clustering Models – K Means and Hierarchical clustering		
Unit:5	**VISUALIZATION**	**6 hours**
Data visualization: Box plot, histogram, scatter plot, heat map – Working with Tableau – Outlier detection – Data Balancing		
Unit:6	**PERFORMANCE EVALUATION in R**	**4 hours**

Syllabus

Loss Function and Error: Mean Squared Error, Root Mean Squared Error – Model Selection and Evaluation criteria: Accuracy, Precision, F1 score, Recall Score – Binary Predictive Classification – Sensitivity – Specificity.

Text Book(s)

1. Ethem Alpaydin, Introduction to Machine Learning, Fourth Edition, MIT Press, 2020
2. Hadley Wickham, Garrett Grolemund, R for data science : Import, Tidy, Transform, Visualize, And Model Data Paperback, 2017

Reference Books

1. Han, J., Kamber, M., Pei, J. Data mining concepts and techniques. Morgan Kaufmann. 2011
2. Carl Shan, Henry Wang, William Chen, Max Song. The Data Science Handbook: Advice and Insight from 25 Amazing Data Scientists. The Data Science Bookshelf. 2016
3. James, G., Witten, D., T., Tibshirani, R. An Introduction to statistical learning with applications in R. Springer. 2013

5CDS-21: R Programming Lab

Credit: 1 **Max. Marks:100 (IA:60, ETE:40)**

0L+0T+2P **End Term Exam: 2 Hours**

SN	List of Experiments
1	● Write R code to create variables and assign values to them. ● Use R operators to perform basic mathematical operations.
2	● Use control flow statements to create conditional and iterative statements.
3	● Write R functions to perform common tasks. ● Import and export data from various sources.
4	Clean and wrangle data to remove errors and inconsistencies.
5	Summarize data using descriptive statistics.
6	Create basic visualizations using R, such as bar charts, line charts, and scatter plots.
7	Create advanced visualizations using R, such as heatmaps, tree maps, and interactive visualizations.
8	Use R to build machine learning models.
9	Use R to create web applications.
10	Use R to work with big data.
11	Use R to deploy models in production.
12	Drawing three dimensional objects and Scenes
13	● Use R to automate tasks and improve efficiency. ● Collaborate with others on data analysis projects using R and Power BI.

B.Tech.: Computer Science & Data Science

5CDS-22: Compiler Design Lab

Credit: 1 **Max. Marks: 100 (IA:60, ETE:40)**

0L+0T+2P **End Term Exam: 2 Hours**

SN	List of Experiments
1	Introduction: Objective, scope and outcome of the course.
2	To identify whether given string is keyword or not.
3	Count total no. of keywords in a file. [Taking file from user]
4	Count total no of operators in a file. [Taking file from user]
5	Count total occurrence of each character in a given file. [Taking file from user]
6	Write a C program to insert, delete and display the entries in Symbol Table.
7	Write a LEX program to identify following: 1. Valid mobile number 2. Valid url 3. Valid identifier 4. Valid date (dd/mm/yyyy) 5. Valid time (hh:mm:ss)
8	Write a lex program to count blank spaces,words,lines in a given file.
9	Write a lex program to count the no. of vowels and consonants in a C file.
10	Write a YACC program to recognize strings aaab,abbb using a^nb^n, where b>=0.
11	Write a YACC program to evaluate an arithmetic expression involving operators +,-,* and /.
12	Write a YACC program to check validity of a strings abcd,aabbcd using grammar a^nb^nc^md^m, where n , m>0
13	Write a C program to find first of any grammar.

RAJASTHAN TECHNICAL UNIVERSITY, KOTA

Syllabus

B.Tech.: Computer Science & Data Science

5CDS-23: Analysis of Algorithms Lab

Credit: 1 **Max. Marks: 100 (IA:60, ETE:40)**

0L+0T+2P **End Term Exam: 2 Hours**

SN	List of Experiments
1	Sort a given set of elements using the Quicksort method and determine the time required to sort the elements. Repeat the experiment for different values of n, the number of elements in the list to be sorted and plot a graph of the time taken versus n. The elements can be read from a file or can be generated using the random number generator.
2	Implement a parallelized Merge Sort algorithm to sort a given set of elements and determine the time required to sort the elements. Repeat the experiment for different values of n, the number of elements in the list to be sorted and plot a graph of the time taken versus n. The elements can be read from a file or can be generated using the random number generator.
3	a. Obtain the Topological ordering of vertices in a given digraph. b. Compute the transitive closure of a given directed graph using Warshall's algorithm.
4	Implement 0/1 Knapsack problem using Dynamic Programming.
5	From a given vertex in a weighted connected graph, find shortest paths to other vertices using Dijkstra's algorithm.
6	Find Minimum Cost Spanning Tree of a given undirected graph using Kruskal's algorithm.
7	a. Print all the nodes reachable from a given starting node in a digraph using BFS method. b. Check whether a given graph is connected or not using DFS method.
8.	Find Minimum Cost Spanning Tree of a given undirected graph using Prim's algorithm.
9.	Implement All-Pairs Shortest Paths Problem using Floyd's algorithm.
10	Implement N Queen's problem using Back Tracking.

B.Tech.: Computer Science & Data Science

5CDS-24: Advance Java Lab

Credit: 1 **Max. Marks: 100 (IA:60, ETE:40)**
L+0T+2P **End Term Exam: 2 Hours**

SN	List of Experiments
1	Introduction To Swing, MVC Architecture, Applets, Applications and Pluggable Look and Feel, Basic swing components : Text Fields, Buttons, Toggle Buttons, Checkboxes, and Radio Buttons
2	Java database Programming, java.sql Package, JDBC driver, Network Programming With java.net Package, Client and Server Programs, Content And Protocol Handlers
3	RMI architecture, RMI registry, Writing distributed application with RMI, Naming services, Naming And Directory Services, Overview of JNDI, Object serialization and Internationalization
4	J2EE architecture, Enterprise application concepts, n-tier application concepts, J2EE platform, HTTP protocol, web application, Web containers and Application servers
5	Server side programming with Java Servlet, HTTP and Servlet, Servlet API, life cycle, configuration and context, Request and Response objects, Session handling and event handling, Introduction to filters with writing simple filter application
6	JSP architecture, JSP page life cycle, JSP elements, Expression Language, Tag Extensions, Tag Extension API, Tag handlers, JSP Fragments, Tag Files, JSTL, Core Tag library, overview of XML Tag library, SQL Tag library and Functions Tag library

Syllabus of

UNDERGRADUATE DEGREE COURSE

B.Tech. VI Semester

Computer Science & Engineering
(Data Science)

Rajasthan Technical University, Kota

Effective from session: 2021-22

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-01: Digital Image Processing

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Introduction to Image Processing:** Digital Image representation, Sampling & Quantization, Steps in image Processing, Image acquisition, color image representation.	04
3	**Image Transformation & Filtering:** Intensity transform functions, histogram processing, Spatial filtering, Fourier transforms and its properties, frequency domain filters, colour models, Pseudo colouring, colour transforms, Basics of Wavelet Transforms.	06
4	**Image Restoration:** Image degradation and restoration process, Noise Models, Noise Filters, degradation function, Inverse Filtering, Homomorphism Filtering.	07
5	**Image Compression:** Coding redundancy, Interpixel redundancy, Psychovisual redundancy, Huffman Coding, Arithmetic coding, Lossy compression techniques, JPEG Compression.	05
6	**Image Segmentation & Representation:** Point, Line and Edge Detection, Thresholding, Edge and Boundary linking, Hough transforms, Region Based Segmentation, Boundary representation, Boundary Descriptors.	05
	Total	28

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-02: Machine Learning

Credit: 3 **Max. Marks: 100(IA:30, ETE:70)**
3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Supervised learning algorithm:** Introduction, types of learning, application, Supervised learning: Linear Regression Model, Naive Bayes classifier Decision Tree, K nearest neighbor, Logistic Regression, Support Vector Machine, Random Forest algorithm	09
3	**Unsupervised learning algorithm:** Grouping unlabelled items using k-means clustering, Hierarchical Clustering, Probabilistic clustering, Association rule mining, Apriori Algorithm, f-p growth algorithm, Gaussian mixture model.	08
4	**Introduction to Statistical Learning Theory**, Feature extraction - Principal component analysis, Singular value decomposition. Feature selection – feature ranking and subset selection, filter, wrapper and embedded methods, Evaluating Machine Learning algorithms and Model Selection.	08
5	**Semi supervised learning, Reinforcement learning:** Markov decision process (MDP), Bellman equations, policy evaluation using Monte Carlo, Policy iteration and Value iteration, Q-Learning, State-Action-Reward-State-Action (SARSA), Model-based Reinforcement Learning.	08
6	**Recommended system,** Collaborative filtering, Content-based filtering Artificial neural network, Perceptron, Multilayer network, Backpropagation, Introduction to Deep learning.	08
	Total	42

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-03: Information Security System

Credit:2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Introduction to security attacks:** services and mechanism, classical encryption techniques- substitution ciphers and transposition ciphers, cryptanalysis, stream and block ciphers.	06
3	**Modern block ciphers**: Block Cipher structure, Data Encryption standard (DES) with example, strength of DES, Design principles of block cipher, AES with structure, its transformation functions, key expansion, example and implementation. Multiple encryption and triple DES, Electronic Code Book, Cipher Block Chaining Mode, Cipher Feedback mode, Output Feedback mode, Counter mode.	06
4	**Public Key Cryptosystems with Applications:** Requirements and Cryptanalysis, RSA cryptosystem, Rabin cryptosystem, Elgamal cryptosystem, Elliptic curve cryptosystem.	06
5	**Cryptographic Hash Functions, their applications:** Simple hash functions, its requirements and security, Hash functions based on Cipher Block Chaining, Secure Hash Algorithm (SHA). Message Authentication Codes, its requirements and security, MACs based on Hash Functions, Macs based on Block Ciphers. Digital Signature, its properties, requirements and security, various digital signature schemes (Elgamal and Schnorr), NIST digital Signature algorithm.	05
6	**Key management and distribution:** symmetric key distribution using symmetric and asymmetric encryptions, distribution of public keys, X.509 certificates, Public key infrastructure. Remote user authentication with symmetric and asymmetric encryption, Kerberos Web Security threats and approaches, SSL architecture and protocol, Transport layer security, HTTPS and SSH.	04
	Total	28

Syllabus
B.Tech.: Computer Science & Data Science
6CDS-04: Computer Architecture and Organization

Credit: 3 **Max. Marks: 100(IA:30, ETE:70)**
3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Computer Data Representation:** Basic computer data types, Complements, Fixed point representation, Register Transfer and Micro-operations: Floating point representation, Register Transfer language, Register Transfer, Bus and Memory Transfers (Tree-State Bus Buffers, Memory Transfer), Arithmetic Micro-Operations, Logic Micro-Operations, Shift Micro-Operations, Arithmetic logical shift unit. Basic Computer Organization and Design Instruction codes, Computer registers, computer instructions, Timing and Control, Instruction cycle, Memory-Reference Instructions, Input-output and interrupt, Complete computer description, Design of Basic computer, design of Accumulator Unit.	10
3	**Programming The Basic Computer:** Introduction, Machine Language, Assembly Language, assembler, Program loops, Programming Arithmetic and logic operations, subroutines, I-O Programming. Micro programmed Control: Control Memory, Address sequencing, Micro program Example, design of control Unit	7
4	**Central Processing Unit:** Introduction, General Register Organization, Stack Organization, Instruction format, Addressing Modes, data transfer and manipulation, Program Control, Reduced Instruction Set Computer (RISC)Pipeline And Vector Processing, Flynn's taxonomy, Parallel Processing, Pipelining, Arithmetic Pipeline, Instruction, Pipeline, RISC Pipeline, Vector Processing, Array Processors	8
5	**Computer Arithmetic:** Introduction, Addition and subtraction, Multiplication Algorithms (Booth Multiplication Algorithm), Division Algorithms, Floating Point Arithmetic operations, Decimal Arithmetic Unit. Input-Output Organization, Input-Output Interface, Asynchronous Data Transfer, Modes Of Transfer, Priority Interrupt, DMA, Input-Output Processor (IOP), CPUIOP Communication, Serial communication.	8
6	**Memory Organization:** Memory Hierarchy, Main Memory, Auxiliary Memory, Associative Memory, Cache Memory, Virtual Memory. Multipreocessors: Characteristics of Multiprocessors, Interconnection Structures, Inter-processor Arbitration, Inter-processor Communication and Synchronization, Cache Coherence, Shared Memory Multiprocessors.	8
	Total	42

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-05: Principles of Artificial Intelligence

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Introduction to AI and Intelligent agent:** Different Approach of AI, Problem Solving : Solving Problems by Searching, Uninformed search, BFS, DFS, Iterative deepening, Bi directional search, Hill climbing, Informed search techniques: heuristic, Greedy search, A* search, AO* search, constraint satisfaction problems.	03
3	**Game Playing:** Minimax, alpha-beta pruning, jug problem, chess problem, tiles problem	06
4	**Knowledge and Reasoning:** Building a Knowledge Base: Propositional logic, first order logic, situation calculus. Theorem Proving in First Order Logic. Planning, partial order planning. Uncertain Knowledge and Reasoning, Probabilities, Bayesian Networks.	06
5	**Learning:** Overview of different forms of learning, Supervised base learning: Learning Decision Trees, SVM, Unsupervised based learning, Market Basket Analysis, Neural Networks.	07
6	**Introduction to Natural Language Processing:** Different issues involved in NLP, Expert System, Robotics.	05
	Total	28

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-06: Cloud Computing

Credit: 3 **Max. Marks: 100(IA:30, ETE:70)**
3L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	01
2	**Introduction:** Objective, scope and outcome of the course. Introduction Cloud Computing: Nutshell of cloud computing, Enabling Technology, Historical development, Vision, feature Characteristics and components of Cloud Computing. Challenges, Risks and Approaches of Migration into Cloud. Ethical Issue in Cloud Computing, Evaluating the Cloud's Business Impact and economics, Future of the cloud. Networking Support for Cloud Computing. Ubiquitous Cloud and the Internet of Things	06
3	**Cloud Computing Architecture:** Cloud Reference Model, Layer and Types of Clouds, Services models, Data centre Design and interconnection Network, Architectural design of Compute and Storage Clouds. Cloud Programming and Software: Fractures of cloud programming, Parallel and distributed programming paradigms-Map Reduce, Hadoop, High level Language for Cloud. Programming of Google App engine.	10
4	**Virtualization Technology:** Definition, Understanding and Benefits of Virtualization. Implementation Level of Virtualization, Virtualization Structure/Tools and Mechanisms, Hypervisor VMware, KVM, Xen. Virtualization: of CPU, Memory, I/O Devices, Virtual Cluster and Resources Management, Virtualization of Server, Desktop, Network, and Virtualization of data-centre.	10
5	**Securing the Cloud:** Cloud Information security fundamentals, Cloud security services, Design principles, Policy Implementation, Cloud Computing Security Challenges, Cloud Computing Security Architecture. Legal issues in cloud Computing. Data Security in Cloud: Business Continuity and Disaster Recovery , Risk Mitigation , Understanding and Identification of Threats in Cloud, SLA-Service Level Agreements, Trust Management	08
6	**Cloud Platforms in Industry:** Amazon web services , Google AppEngine, Microsoft Azure Design, Aneka: Cloud Application Platform -Integration of Private and Public Clouds Cloud applications: Protein structure prediction, Data Analysis, Satellite Image Processing, CRM	07
	Total	42

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-11: Artificial Neural Network

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

Course Objectives:

1. To understand the biological neural network and to model equivalent neuron models.
2. To understand the architecture, learning algorithm and issues of various feed forward and feedback neural networks.

Course Outcomes: By completing this course the student will be able to:

- Create different neural networks of various architectures both feed forward and feed backward.
- Perform the training of neural networks using various learning rules.
- Perform the testing of neural networks and do the perform analysis of these networks for various pattern recognition applications.

Detailed Syllabus: (per session plan)

UNIT	Contents
1	Introduction: A Neural Network, Human Brain, Models of a Neuron, Neural Networks viewed as Directed Graphs, Network Architectures, Knowledge Representation, Artificial Intelligence and Neural Networks. Learning Process: Error Correction Learning, Memory Based Learning, Hebbian Learning, Competitive, Boltzmann Learning, Credit Assignment Problem, Memory, Adaption, Statistical Nature of the Learning Process.
2	Single Layer Perceptrons: Adaptive Filtering Problem, Unconstrained Organization Techniques, Linear Least Square Filters, Least Mean Square Algorithm, Learning Curves, Learning Rate Annealing Techniques, Perceptron –Convergence Theorem, Relation Between Perceptron and Bayes Classifier for a Gaussian Environment. Multilayer Perceptron: Back Propagation Algorithm XOR Problem, Heuristics, Output Representation and Decision Rule, Computer Experiment, Feature Detection.
3	Back Propagation: Back Propagation and Differentiation, Hessian Matrix, Generalization, Cross Validation, Network Pruning Techniques, Virtues and Limitations of Back Propagation Learning, Accelerated Convergence, Supervised Learning.
4	Self-Organizing Maps (SOM): Two Basic Feature Mapping Models, Self-Organization Map, SOM Algorithm, Properties of Feature Map, Computer Simulations, Learning Vector Quantization, Adaptive Pattern Classification.
5	Neuro Dynamics: Dynamical Systems, Stability of Equilibrium States, Attractors, Neuro Dynamical Models, Manipulation of Attractors as a Recurrent Network Paradigm.

	Hopfield Models – Hopfield Models, Computer Experiment.

TEXT BOOKS:

1. Neural Networks a Comprehensive Foundations, Simon Haykin, PHI edition.

REFERENCE BOOKS:

1. Artificial Neural Networks - B. Vegnanarayana Prentice Hall of India P Ltd 2005
2. Neural Networks in Computer Intelligence, Li Min Fu MC GRHILL EDUCATION 2003:
3. Neural Networks -James A Freeman David M S Kapura Pearson Education 2004.
4. Introduction to Artificial Neural Systems Jacek M. Zurada, JAICO Publishing House Ed. 2006.

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-12: Nature Inspired Computing (NLP)

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

Course Objectives:

- To establish basic knowledge in NP hard problems and understand the need for approximation algorithms.
- Design algorithms that include operators, representations, fitness functions and potential hybridizations for non-trivial problems.
- Design algorithms that utilize the collective intelligence of simple organisms to solve problems.
- Design and implement an artificial neural network that employs learning to solve nontrivial problems.

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

1. Understand fundamental concepts of NP-hardness and computational complexity
2. Understand the strengths, weaknesses and appropriateness of nature-inspired algorithms.
3. Apply nature-inspired algorithms to optimization, design and learning problems.
4. Analyse the Behaviour systems of nature inspired algorithm applied in real world problems.
5. Understand the theory behind the design of immune networks and DNA computing and their potential applications.

Detailed Syllabus: (per session plan)

UNIT	Contents
1	Evolutionary Systems: Pillars of Evolutionary Theory, The Genotype, Artificial Evolution, Genetic representations, Initial Population, Fitness Functions, Selection and Reproduction, Genetic Operators, Evolutionary Measures, Types of Evolutionary Algorithms
2	Collective Systems: Particle Swarm Optimization Algorithm, Hybrid PSO algorithms, Ant Colony Optimization, Artificial Bee Colony, Firefly Algorithm Artificial Neural Networks: History, Mathematical model of neuron, ANN architectures, learning rules Backpropagation network, Backpropagation learning and its applications, Variants of BPA.
3	Behaviour in Cognitive Science, Behaviour in Artificial Intelligence, Behaviour-Based Robotics, Biological Inspiration for Robots, Robots as Biological Models, Robot Learning, Evolution of Behavioural Systems Evolution and Learning in Behavioural Systems, Evolution and Neural Development in Behavioural Systems.
4	Immuno Computing: Introduction- Immune System, Physiology and main components,Immune Network Theory- Danger Theory, Evaluation Interaction-Immune Algorithms, Bone Marrow Models, Forest's Algorithm, Artificial Immune Networks.

5	DNA Computing: Motivation, DNA Molecule, Adleman's experiment, Test tube programming language, Universal DNA Computers , PAM Model , Splicing Systems,Lipton's Solution to SAT Problem , Scope of DNA Computing , From Classical to DNA Computing.

TEXT BOOKS:

1. Xin-She Yang, “Nature-Inspired Computation and Swarm Intelligence Algorithms, Theory and Applications”, Elsevier, Academic Press, 2020.

REFERENCE BOOKS:

1. Leandro Nunes de Castro, "Fundamentals of Natural Computing, Basic Concepts, Algorithms and Applications", Chapman & Hall/ CRC, Taylor and Francis Group, 2007.
2. Floreano D. and Mattiussi C., "Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies", MIT Press, Cambridge, MA, 2008.
3. Licheng Jiao, Ronghua Shang, Fang Liu, Weitong Zhang , Brain and Nature-Inspired Learning, Computation and Recognition, Elsevier, 2020.

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-13:Big Data Analytics & Hadoop

Credit: 2 **Max. Marks: 100(IA:30, ETE:70)**
2L+0T+0P **End Term Exam: 3 Hours**

Course Objectives:

- To establish basic knowledge in NP hard problems and understand the need for approximation algorithms.
- Design algorithms that include operators, representations, fitness functions and potential hybridizations for non-trivial problems.
- Design algorithms that utilize the collective intelligence of simple organisms to solve problems.
- Design and implement an artificial neural network that employs learning to solve non- trivial problems.

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

1. Understand fundamental concepts of NP-hardness and computational complexity
2. Understand the strengths, weaknesses and appropriateness of nature-inspired algorithms.
3. Apply nature-inspired algorithms to optimization, design and learning problems.
4. Analyze the Behaviour systems of nature inspired algorithm applied in real world problems.
5. Understand the theory behind the design of immune networks and DNA computing and their potential applications.

Detailed Syllabus: (per session plan)

UNIT	Contents
1	Introduction to Big Data and Hadoop : - Characteristics of big data and its impact on industries. - Overview of Hadoop history, ecosystem components. - Hadoop Architecture: HDFS structure, basic data operations. - Setting up a Hadoop cluster: Installation, configuration basics.
2	Hadoop Data Storage and Processing : - Understanding HDFS: Storage model, fault tolerance. - Introduction to MapReduce: Concepts, basic data flow. - MapReduce Workflow: Mapper, Reducer, Combiner, basic use cases. - Hands-on MapReduce: Writing simple MapReduce jobs, analyzing output.
3	Hadoop Ecosystem and Data Processing: - Hadoop Ecosystem Overview: HBase, Hive, Pig, Spark. - Hive Basics: Creating tables, querying structured data. - Pig Fundamentals: Data flow, basic Pig Latin scripting. - Introduction to Spark: RDDs, basic transformations and actions.
4	Big Data Analytics Techniques:- Introduction to Big Data Analytics: Role, challenges. - Data Preprocessing: Basics of cleaning, transformation, feature extraction. - Exploratory Data Analysis (EDA): Using visualizations

	for insights. - Introduction to Machine Learning on Big Data: Overview of algorithms.
5	Advanced Topics and Capstone Project:- Advanced Hadoop Concepts: Brief on YARN, security mechanisms. - Real-time Data Processing: Basic introduction to Kafka and Storm. - Big Data Analytics with Spark: Overview of MLlib, GraphX. - Capstone Project: Applying learned techniques to a real-world problem.

TEXT BOOKS:

1. White, T. (2015). Hadoop: The Definitive Guide. O'Reilly Media.

REFERENCE BOOKS:

1. “Hadoop: The Definitive Guide” by Tom White.
2. “Big Data Analytics with R and Hadoop” by Vignesh Prajapati.
3. “Learning Spark” by Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia.
4. “Big Data for Dummies” by Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman.

Syllabus
B.Tech.: Computer Science & Data Science

6CDS-21: Digital Image Processing Lab

Credit: 1.5 **Max. Marks: 100(IA:60, ETE:40)**
0L+0T+3P **End Term Exam: 2 Hours**

SN	List of Experiments
1	Point-to-point transformation. This laboratory experiment provides for thresholding an image and the evaluation of its histogram. Histogram equalization. This experiment illustrates the relationship among the intensities (gray levels) of an image and its histogram.
2	Geometric transformations. This experiment shows image rotation, scaling, and translation. Two-dimensional Fourier transform
3	Linear filtering using convolution. Highly selective filters.
4	Ideal filters in the frequency domain. Non Linear filtering using convolutional masks. Edge detection. This experiment enables students to understand the concept of edge detectors and their operation in noisy images.
5	Morphological operations: This experiment is intended so students can appreciate the effect of morphological operations using a small structuring element on simple binary images. The operations that can be performed are erosion, dilation, opening, closing, open-close, close-open.

Syllabus
B.Tech.: Computer Science & Data Science
6CDS-22: Machine Learning Lab

Credit: 1.5 **Max. Marks: 100(IA:60, ETE:40)**
0L+0T+3P **End Term Exam: 2 Hours**

SN	List of Experiments
1	Implement and demonstrate the FIND-Salgorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.
2	For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithmto output a description of the set of all hypotheses consistent with the training examples.
3	Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge toclassify a new sample
4	Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets
5	Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
6	Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.
7	Write a program to construct aBayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.
8	Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
9	Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
10	Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Syllabus

B.Tech.: Computer Science & Data Science

6CDS-23: Python Lab

Credit: 1.5 **Max. Marks: 100(IA:60, ETE:40)**

0L+0T+3P **End Term Exam: 2 Hours**

SN	List of Experiments
1	Write a program to demonstrate basic data type in python.
2	Write a program to compute distance between two points taking input from the user Write a program add.py that takes 2 numbers as command line arguments and prints its sum.
3	Write a Program for checking whether the given number is an even number or not. Using a for loop, write a program that prints out the decimal equivalents of 1/2, 1/3, 1/4, . . . , 1/10
4	Write a Program to demonstrate list and tuple in python. Write a program using a for loop that loops over a sequence. Write a program using a while loop that asks the user for a number, and prints a countdown from that number to zero.
5	Find the sum of all the primes below two million. By considering the terms in the Fibonacci sequence whose values do not exceed four million, WAP to find the sum of the even-valued terms.
6	Write a program to count the numbers of characters in the string and store them in a dictionary data structure Write a program to use split and join methods in the string and trace a birthday of a person with a dictionary data structure
7	Write a program to count the frequency of characters in a given file. Can you use character frequency to tell whether the given file is a Python program file, C program file or a text file? Write a program to count the frequency of characters in a given file. Can you use character frequency to tell whether the given file is a Python program file, C program file or a text file?
8	Write a program to print each line of a file in reverse order. Write a program to compute the number of characters, words and lines in a file.
9	Write a function nearly equal to test whether two strings are nearly equal. Two strings a and b are nearly equal when a can be generated by a single mutation on. Write a function to compute gcd, lcm of two numbers. Each function shouldn't exceed one line.
10	Write a program to implement Merge sort. Write a program to implement Selection sort, Insertion sort.

Syllabus
B.Tech.: Computer Science & Data Science

6AID4-24/6CAI4-24: Mobile Application Development Lab

Credit: 1.5 **Max. Marks: 100(IA:60, ETE:40)**
0L+0T+3P **End Term Exam: 2 Hours**

SN	List of Experiments
1	To study Android Studio and android studio installation. Create a "Hello World" application.
2	To understand Activity, Intent, Create sample application with login module.(Check username and password).
3	Design simple GUI application with activity and intents e.g. calculator.
4	Develop an application that makes use of RSS Feed.
5	Write an application that draws basic graphical primitives on the screen
6	Create an android app for database creation using SQLite Database.
7	Develop a native application that uses GPS location information
8	Implement an application that writes data to the SD card.
9	Design a gaming application
10	Create an application to handle images and videos according to size.