Gurugram University, Gurugram

GENERAL COURSE STRUCTURE AND CREDIT DISTRIBUTION

CREDIT DISTRIBUTION

POSTGRADUATE PROGRAMME

Scheme of Programme Master of Science in Computer Science (AI & DS).

(Scheme PG A1: Postgraduate Programmes (Course work only)

Semester 1

Course Code	Course Title	Course ID	L (Hrs)	T (Hrs)	P (Hrs)	L Credits	T Credits	P Credits	Total Credits	MARKS TI	MARKS TE	MARKS PI	MARKS PE	MARKS Total
	Core Course(s)													
CC-A01	Data Structures & Algorithms	241/CSAI/CC101	3	0	2	3	0	1	4	25	50	5	20	100
CC-A02	Database Management System	241/CSAI/CC102	3	0	2	3	0	1	4	25	50	5	20	100
CC-A03	Fundamentals of Data Science	241/CSAI/CC103	3	0	2	3	0	1	4	25	50	5	20	100
	Discipline Specific Elective Courses													
DSE-01	Probability and Statistics	241/CSAI/DSE101	2	1	0	2	1	0	3	25	50	-	-	75
	Multidisciplinary Course(s)													
MDC-01	One From Pool	241/CSAI/MD101	2	1	-	2	1	-	3	25	50	-	-	75
	Ability Enhancement Course(s)													
AEC-01	One From Pool	241/CSAI/AE101							2					50
	Value-added Course(s)													
VAC-01	One From Pool	241/CSAI/VA101	2	-	-	2	-	-	2	15	35	-	-	50
Total Credits									**22**					**550**

Semester 2

Course Code	Course Title	Course ID	L (Hrs)	T (Hrs)	P (Hrs)	L Credits	T Credits	P Credits	Total Credits	MARKS TI	MARKS TE	MARKS PI	MARKS PE	MARKS Total
Core Course(s)														
CC-A04	Data Analytics	241/CSAI/CC201	3	0	2	3	0	1	4	25	50	5	20	100
CC-A05	Data Mining	241/CSAI/CC202	3	0	2	3	0	1	4	25	50	5	20	100
CC-A06	Artificial Intelligence	241/CSAI/CC203	3	0	2	3	0	1	4	25	50	5	20	100
Discipline Specific Elective Courses														
DSE-02	Information Retrieval Systems	241/CSAI/DSE201	2	1	0	2	1	0	3	25	50	-	-	75
Multidisciplinary Course(s)														
MDC-02	One From Pool	241/CSAI/MD201	2	1	-	2	1	-	3	25	50	-	-	75
Ability Enhancement Course(s)														

AEC-02	One From Pool	241/CSAI/AE201							2					50
Skill Enhancement Course(s)														
SEC-01	One From Pool	241/CSAI/SE201	2	-	-	2	-	-	2	15	35	-	-	50
Total Credits									**22**					**550**

Semester 3

Course Code	Course Title	Course ID	L	T	P	L	T	P	Total Credits	MARKS				
			(Hrs)			Credits				TI	TE	PI	PE	Total
Core Course(s)														
CC-A07	Design and Analysis of Algorithms	241/CSAI/CC301	3	0	2	3	0	1	4	25	50	5	20	100
CC-A08	Machine Learning – I	241/CSAI/CC302	3	0	2	3	0	1	4	25	50	5	20	100
CC-A09	Fuzzy Systems and Applications	241/CSAI/CC303	3	0	2	3	0	1	4	25	50	5	20	100
Discipline Specific Elective Courses														
DSE-03	Internet of Things	241/CSAI/DSE301	2	1	0	2	1	0	3	25	50	-	-	75
Multidisciplinary Course(s)														
MDC-03	One From Pool	241/CSAI/MD301	2	1	-	2	1	-	3	25	50	-	-	75
Skill Enhancement Course(s)														
SEC-02	One From Pool	241/CSAI/SE301	2	-	-	2	-	-	2	15	35	-	-	50
Value-added Course(s)														
VAC-02	One From Pool	241/CSAI/VA301	2	-	-	2	-	-	2	15	35	-	-	50
Seminar														
Seminar	Seminar		2	-	-	2	-	-	2					50
Internship/Field Activity#														
Internship/Field Activity	Internship/Field Activity		-	-	8	-	-	4	4					100
Total Credits									**28**					

#Four credits of internship earned by a student during summer internship after 2nd semester will be counted in 3rd semester of a student who pursue 2 year PG Programme without taking exit option.

Semester 4

Course Code	Course Title	Course ID	L	T	P	L	T	P	Total Credits	MARKS				
			(Hrs)			Credits				TI	TE	PI	PE	Total
Core Course(s)														
CC-A10	Big Data Analytics	241/CSAI/CC401	3	0	2	3	0	1	4	25	50	5	20	100
CC-A11	Deep Learning	241/CSAI/CC402	3	0	2	3	0	1	4	25	50	5	20	100
Discipline Specific Elective Courses														
DSE-04	Computational Intelligence	241/CSAI/DSE401	2	1	0	2	1	0	3	25	50	-	-	75
Multidisciplinary Course(s)														
MDC-04	One From Pool	241/CSAI/MD401	2	1	-	2	1	-	3	25	50	-	-	75
Ability Enhancement Course(s)														
AEC-03	One From Pool	241/CSAI/AE401							2	15	35	-	-	50
Community Engagement/Field Work/Survey/Seminar/Project/Training														
Seminar	**Seminar**		-	-	12	-	-	6	6					150
Total Credits									**22**					

Semester 1

DATA STRUCTURES & ALGORITHMS

Course Id - 241/CSAI/CC101

Semester	1			
Course code	CC-A01			
Category	Core Course			
Course title	Data Structures & Algorithms			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. Describe types of data structures and operations that can be implemented on these data structures.
2. Demonstrate the use of various data structure and their related operations.
3. Apply appropriate data structures with respect to effective storage of data and efficiency of the required operations on data for solving real world problems.
4. Analyse the algorithms. time complexity of searching and algorithms.
5. Formulate data structures and prescribe operations for given real world situations

UNIT - I

Introduction: Fundamentals of pointers in C, pointer declaration, passing pointer to functions, pointers and arrays, dynamic memory allocation, Definition of Algorithm, Data Abstraction, Performance Analysis & Measurement, Files and related operations in C. Data Structures vs Data Types.

Searching and Sorting Techniques: Searching techniques: Linear and Binary search, Sorting

techniques: Selection, Bubble, Insertion, Merge sort, Quicksort, List and Table Sorting

UNIT - II

Linear Data Structures- I Arrays: Definition of array, Array storage, sparse arrays; Transpose, addition, and multiplication of sparse matrices, Stacks and Queues and their applications, expression evaluation, A mazing problem; multiple stacks and queues in an array, Application of stacks recursion polish expression and their compilation conversion of infix expression to prefix and postfix expression, Tower of Hanoi problem.

UNIT - III

Linear Data Structures- II Linked Lists; definition, allocation for stacks and queues. Examples of linked lists, polynomial addition, comparison of sequential and linked allocation of storage; inversion, concatenation & copying of the lists. Implementations in C language. Doubly Linked List: Definition of circular and doubly linked list, header node, insertion and deletion, sparse matrix, representation using doubly linked lists. Examples for application of doubly linked lists; dynamic storage management; node structures, routines for allocation and deallocation, generalized lists and recursive algorithms for copying and comparison of lists.

UNIT - IV

Non Linear Data Structures Trees, Basic concepts and definitions of a tree and binary tree and associated terminology, Binary tree traversal techniques, Binary tree representation of trees, transformation of trees into binary trees, some more operations on binary trees, Binary Search Trees, Heaps and heapsort, threaded binary trees, Graphs: Representation of graphs and their traversal, Minimum cost Spanning Trees.

TEXT AND REFERENCE BOOKS:

1. Seymour Lipschutz: Data Structures with C, Schaum's outline by TMH
2. E Horowitz and S. Sahni: Fundamentals of Data Structures in C, Second
3. Edition, Universities Press, Hyderabad.
4. R.B. Patel: Expert Data Structures in C, Khanna Publishers,2001.
5. R.L. Kruse: Data Structures & Program Design in C, PHI.
6. D.F. Knuth: The art of Computer Programming Vol 1, Narosa Publications,1985.
7. Byron S. Gottfried & J K Chhabra: Theory and Problems of Programming with C
8. Language, Schaum Series, TMH,2005.

DATA STRUCTURES & ALGORITHMS LAB

List of Subject related Experiments:

1. Write a program to search an element in a two-dimensional array using linear search.
2. Using iteration & recursion concepts write programs for finding the element in the array using Binary Search Method
3. Write a program to perform following operations on tables using functions only
   (a) Addition (b) Subtraction (c) Multiplication (d) Transpose
4. Using iteration & recursion concepts write the programs for Quick Sort Technique
5. Write a program to implement the various operations on string such as length of string concatenation, reverse of a string & copy of a string to another.
6. Write a program for swapping of two numbers using ‘call by value’ and ‘call by reference strategies.
7. Write a program to implement binary search tree.
8. (Insertion and Deletion in Binary search Tree)
9. Write a program to create a linked list & perform operations such as insert, delete, update, reverse in the link list
10. Write the program for implementation of a file and performing operations such as insert, delete, update a record in the file.
11. Create a linked list and perform the following operations on it
    (a) add a node (b) Delete a node

DATABASE MANAGEMENT SYSTEM

Course Id - 241/CSAI/CC102

Semester	1			
Course code	CC-A02			
Category	Core Course			
Course title	Database Management System			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. At the end of this course, students will demonstrate the ability to
2. Understand the database concepts and structures.
3. Understand data modeling and database development process.
4. Construct and normalize conceptual data models. Implement a relational database into a database management system.
5. CO4 Use database management systems (Oracle SQL Plus).
6. CO5 Become proficient in using database query language (SQL)

UNIT - I

Introduction: Characteristics of database approach, data models, DBMS architecture and data independence.

E-R Modeling: Entity types, Entity set, attribute and key, Relationships, Relation types, Roles and Structural constraints, Weak entities, Enhanced ER Model.

Database Languages: DDL, DML, Database Access for applications Programs, Database Users and

Administrator, Transaction Management, Database system Structure, Storage Manager, Query Processor.

UNIT - II

Relational Model: Introduction to the Relational Model, Integrity Constraint over Relations, Enforcing Integrity constraints, Querying relational data, Logical data base Design, Introduction to views, Destroying/altering Tables and Views.

Relational Algebra and Calculus: Relational Algebra, Set operations, Selection and projection, renaming, Joins, Division, Examples of Algebra overviews, Relational calculus: Tuple relational Calculus, Domain relational calculus, Expressive Power of Algebra and Calculus.

UNIT - III

Schema Refinement, Functional dependencies: Problems Caused by redundancy, Decompositions, Problem related to decomposition, Normalization : FIRST, SECOND, THIRD Normal forms, BCNF, Lossless join Decomposition, Dependency preserving Decomposition, Schema refinement in Data base Design, Multi valued Dependencies, forth Normal Form.

Transaction Management: ACID Properties, Transactions and Schedules, Concurrent Execution of transaction, Serializability and recoverability.

UNIT - IV

Concurrency Control: Introduction to Lock Management, Lock Conversions, Dealing with Dead Locks, Concurrency without Locking, Recovery Techniques, Database Security.

Introduction to Oracle : Getting started, Modules of Oracle, Invoking SQLPLUS, Data types, Data Constraints, Operators, Data manipulation - Create, Modify, Insert, Delete and Update; Searching, Matching and Oracle Functions.

SQL* Forms: Basic concepts, Form Construction, Creating default form, user-defined form, multiple-record form, Master-detail form.

TEXT AND REFERENCE BOOKS:

1. Raghurama Krishnan: Data base Management Systems, Johannes Gehrke, Tata McGraw Hill.
2. Siberschatz, Korth : Data base System Concepts, McGraw Hill.
3. P. Radha Krishna : Database Management Systems, HI-TECH Publications.
4. C.J. Date : Introduction to Database Systems, Pearson Education.
5. Rob & Coronel : Data base Systems design, Implementation, and Management,Thomson.
6. Elmasri Navrate : Data base Management System, Pearson Education.
7. Mathew Leon : Data base Management System, Leon Vikas Publishers.

DATABASE MANAGEMENT SYSTEM LAB

List of lab experiments

1. Creation of a database and writing SQL queries to retrieve information from the database.
2. Performing Insertion, Deletion, Modifying, Altering, Updating and Viewing records based on conditions.
3. Creation of Views, synonyms, Sequences, indexes, and save points.
4. Creating an employee database to set various constraints.
5. Creating relationships between the databases. iv. Study of PL/SQL block.
6. Write a PL/SQL block to satisfy some conditions by accepting input from the user.
7. Write a PL/SQL block that handles all types of exceptions.
8. Creation of Procedures
9. Creation of database triggers and functions
10. Mini project (Application Development using Oracle/ MySQL)
    a. Inventory Control System
    b. Material Requirement Processing.
    c. Hospital Management System.
    d. Railway Reservation System.
    e. Personal Information System.
    f. Web-Based User Identification System.
    g. Time Table Management System.
    h. Hotel Management

FUNDAMENTALS OF DATA SCIENCE
Course Id - 241/CSAI/CC103

Semester	1			
Course code	CC-A03			
Category	Core Course			
Course title	Fundamentals of Data Science			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE Outcomes:

1. Understand and implement the basics of programming in Python
2. Apply the Numpy package for numerical calculations in Python
3. Apply Pandas package for loading and preprocessing data in Python. Implement various data visualization tools of Python on real world data.
4. Understand and implement the Machine Learning Concepts in Python.

UNIT - I

Overview of Python Programming Concepts: The concept of data types, variables, assignments, numerical types, operators and expressions, Control Structures, String manipulations, File Handling – creating, reading/writing text/number files, Dictionaries, Functions, OOPs Concepts.

UNIT - II

Introduction to Numpy - Creation on Array, Array generation from Uniform distribution, Random

array generation, reshaping, maximum and minimum, reshaping, Arithmetic operations, Mathematical functions, Bracket Indexing and Selection, Broadcasting, Indexing a 2D array (matrices); Data Manipulation with Pandas -Creating a Series - from lists, arrays and dictionaries, Storing data in series from intrinsic sources, Creating Data Frames, Imputation, Grouping and aggregation, Merging, Joining, Concatenation, Find Null Values or Check for Null Values, Reading data from csv, txt, excel, web.

UNIT - III

Introduction to Visualization - Installing and setting up visualization libraries, Canvas and Axes, Subplots, Common plots – scatter, histogram, boxplot, Logarithmic scale, Placement of ticks and custom tick labels, Pandas Viz, Style Sheets, Plot type, Area, Barplots, Histograms, Line Plots, Scatter Plots, BoxPlots, Hexagonal Bin Plot, Distribution Plots, Categorical Data Plots, Combining Categorical Plots, Matrix Plots, Regression Plots, Grids; Python Visualizations toolkits/libraries.

UNIT - IV

Introduction to Machine Learning with SciKit-Learn & PyTorch– Data Representation and basic functions Estimator, parameters & model validation, Model Selection, Curve, Grid search, Feature engineering, Naive Bayes Classification, Linear regression, Introduction to NLP with NLTK and its functions, modules like speech tagging, tokenization, parsing, segmentation, recognition, cleaning & normalization of text.

TEXT AND REFERENCE BOOKS:

1. Charles Dierbach., Introduction to Python using Computer Science, Wiley Publications, Second Edition, 2015
2. Mark Lutz , Learning Python, O'Reilly publications , Fifth Edition, 2015
3. Jake Vandar Plas, Python Data Science Handbook, O'Reilly , 2016
4. Paul Barry, Head First Python, Orielly Publications, Second Edition, 2010

Fundamentals of Data Science Lab

List of Experiments:

1. Python program to display details about the operating system, working directory, files And directories in the current directory, lists the files and all directories, scan and classify them as directories and files.
2. Python program to convert an array to an array of machine values and vice versa
3. Python program to get information about the file pertaining to the file mode and to get time values with components using local time and gm time.

4. Python program to connect to Google using socket programming
5. Python program to perform Array operations using Numpy package
6. Python program to perform Data Manipulation operations using Pandas package.
7. Python program to display multiple types of charts using Matplotlib package
8. Python program to perform File Operation on Excel Data Set
9. Python program to implement with Python Sci Kit-Learn & NLTK.
10. Python program to implement with Python NLTK/Spicy/Py NLPI.

PROBABILITY AND STATISTICS

Course Id - 241/CSAI/DS101

Semester	1			
Course code	DSE-01			
Category	Discipline Specific Elective Course			
Course title	Probability and Statistics			
Scheme and Credits	L	T	P	Credits
	2	1	0	3
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	-			
Practical External	-			
Total	75 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit..*

COURSE Outcomes:

1. Analyze the data using different descriptive measures and present graphically
2. Compute the probabilities of events along with an understanding of the random variables, expectation, variance and distributions.
3. Understand the estimation of mean and variance and their respective one-sample and ksample hypothesis tests.
4. Understand the law of large numbers and the central limit theorem and how these concepts are used to model various random phenomena

UNIT - I

Probability: Classical, relative frequency and axiomatic definitions of probability, addition rule and conditional probability, multiplication rule, total probability, Bayes' Theorem and independence, problems.

Descriptive Statistics: Graphical representation, measures of locations and variability.

UNIT - II

Random Variables: Discrete, continuous and mixed random variables, probability mass, probability density and cumulative distribution functions, mathematical expectation, Function of a random variable, problems. correlation and regression, independence of random variable

UNIT - III

Special Distributions: Discrete uniform, binomial, geometric, Poisson, continuous uniform, exponential, gamma, normal, problems.

Sampling Distributions: The Central Limit Theorem, distributions of the sample mean and the sample variance for a normal population, Chi-Square, t and F distributions, problems

UNIT - IV

Testing of Hypotheses: Null and alternative hypotheses, the critical and acceptance regions, two types of error, power of the test, the most powerful test and, tests for one sample and two sample problems for normal populations, Chisquare goodness of fit test and its applications, problems

TEXT AND REFERENCE BOOKS:

1. An Introduction to Probability and Statistics by V.K. Rohatgi & A.K. Md. E. Saleh, Wiley, (2008), 3rd ed.
2. Introduction to Probability and Statistics by J.S. Milton & J.C. Arnold, Boston, London, McGraw-Hill, (2006), 4th ed.
3. Introduction to Probability Theory and Statistical Inference by H.J. Larson, John Wiley & Sons, (2005) 3rd ed.
4. Introduction to Probability and Statistics for Engineers and Scientists by S.M. Ross, Elsevier, (2014), 4th ed.

Semester 2

DATA ANALYTICS

Semester	2			
Course code	CC-A04			
Category	Core Course			
Course title	Data Analytics			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5			
Practical External	20			
Total	100 marks			
Duration of Exam	3 hrs			

***Note**: The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. Understanding the basics of data Analytics
2. Using versatile and flexible languages (Python and R programming) for supporting data science
3. Using data processing for collecting and manipulating the data into the usable and desired form.
4. Using data visualization to easily access the huge amount of data in visuals.
5. Using statistics to collect and analyze the numerical data in a large amount and finding meaningful insights.

UNIT-I

Data Analytics an Introduction: Computer Science, Data Analytics, and Real Science, What is Data Science? Need for Data Analytics, Data Analytics Components, Tools for Data Analytics, Data Analytics Lifecycle, Applications of Data Analytics.

Python Programming for Data Analytics: Introduction to Python Programming (Python

Basics,Python Data Structures, Python Programming Fundamentals, Working with Data in Python, Working with NumPy, Pandas, SciPy, and Matplotlib).

UNIT-II

Data Pre-processing: Need of Data Pre-processing, Data Pre-processing Methods: Data Cleaning, Data Integration, Data Transformation, Data Reduction; Feature Scaling (Normalization and Standardization), Splitting dataset into Training and Testing set.

Statistical Data Analysis: Measuring Central Tendency, Measuring Variance, Normal Distribution, Binomial Distribution, Poisson Distribution, Bernoulli Distribution, P- Value, Correlation, Chi-square Test, Linear Regression

UNIT-III

Classification: Need and Applications of Classification, Logistic Regression, Decision tree, Tree induction algorithm – split algorithm based on information theory, split algorithm based on Gini index; Random forest classification, Naïve Bayes algorithm; K-Nearest Neighbours (K-NN), Evaluating Classification Models' Performance (Sensitivity, Specificity, Precision, Recall, etc).

Clustering: Need and Applications of Clustering, Partitioned methods, Hierarchical methods,

UNIT-IV

Working with text data: types of data represented as strings, representing text data as a bag of words, stop words, rescaling, advanced tokenization, stemming and lemmatization.

Data Visualization: Chart Properties, Chart Styling, Box Plots, Heat Maps, Scatter Plots, Bubble Charts, 3D Charts, Time Series, Geographical Data, Graph Data

TEXT AND REFERENCE BOOKS:

1. Data Science for Dummies by Lillian Pierson and Jake Porway
2. An Introduction to Statistical Learning by Gareth James, Daniela Witten, et al.
3. Introduction to machine learning with python: Andreas C. Muller and Sarah Guido

DATA ANALYTICS LAB

List of lab experiments

1. Descriptive Statistics: Calculate and interpret measures of central tendency, dispersion, and skewness for a dataset.
2. Implement strategies such as mean/median imputation, forward or backward filling, or using machine learning models to predict missing values.
3. Identify outliers using statistical methods (e.g., Z-score, IQR) or machine learning techniques (e.g., isolation forest) and decide whether to remove them or transform them.
4. Linear Regression: Build a simple linear regression model to predict a continuous variable based on one predictor variable.
5. Logistic Regression: Construct a logistic regression model to predict a binary outcome based on predictor variables.
6. Data Visualization: Create various types of plots (e.g., bar charts, histograms, scatter plots) to explore relationships within a dataset.
7. Clustering Analysis: Apply k-means clustering to segment data into distinct groups based on similarities in features.
8. Decision Trees: Build a decision tree model to classify data into categories based on multiple input variables.
9. Text Analysis: Perform sentiment analysis on text data (e.g., customer reviews, social media comments) to understand public opinion.
10. Apply Bayesian inference to update beliefs and make predictions based on prior knowledge and observed data.

DATA MINING

Semester	2			
Course code	CC-A05			
Category	Core Course			
Course title	Data Mining			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5			
Practical External	20			
Total	100 marks			
Duration of Exam	3 hrs			

***Note**: The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. To understand Concept of Data Mining, Data warehousing and schemas for multidimensional Databases.
2. To understand Basic Statistics in order to apply data mining techniques.
3. To analyse the data, identify the problems, and choose the relevant models and algorithms to apply.
4. To combine and consolidate data from various databases scattered throughout a company into a Datawarehouse.
5. To characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering.

UNIT-I

BASIC STATISTICS:

Statistical descriptions of data: mean, median, mode, Measuring dispersion of data: range, quartiles,

variance, standard deviation, chi-square test, Correlation coefficient and covariance, Regression analysis.

UNIT-II

DATA PREPROCESSING:

Introduction to Data preprocessing, Need to preprocess the data, Data cleaning: missing values, Data integration: Redundancy, Data reduction and its strategies, Data transformation and discretization, Strategies of data transformation.

UNIT-III

INTRODUCTION TO DATA MINING AND ITS ALGORITHMS:

Introduction to Data Mining and algorithms, Processes, Market Basket Analysis, The Apriori Algorithm, Decision Tree.

INTRODUCTION TO DATA WAREHOUSING AND DATA WRANGLING:

Data Warehouse, Difference between Operational Database systems and Data Warehouse, From Table and Spreadsheets to Data Cubes, Schemas for Multidimensional Databases: Star, Snowflakes and Fact Constellations, Data Wrangling, Combining and Merging DataSets, Reshaping and Pivoting, Data Transformation, String Manipulation, Regular Expressions (Regex)

UNIT-IV

ETL:

ETL Phase 1 Data Wrangling before the Load, ETL Phase 2 Step-by-step guide to uploading data using SSIS, Handling errors during ETL Phases 1,2, ETL Phase 3 Data Wrangling after the load, Handling errors during ETL Phase 3, Different types of ETL tools.

TEXT AND REFERENCE BOOKS:

1. Jiawei Han, Micheline Kamber and Jian Pei, “Data Mining Concepts and Techniques”, Third Edition, Elsevier, 2011.
2. Alex Berson and Stephen J. Smith “Data Warehousing, Data Mining & OLAP”, Tata McGraw – Hill Edition, Tenth Reprint 2007
3. K.P. Soman, Shyam Diwakar and V. Ajay “Insight into Data mining Theory and Practice”, Easter Economy Edition, Prentice Hall of India, 2006
4. G. K. Gupta “Introduction to Data Mining with Case Studies”, Easter Economy Edition, Prentice Hall of India, 2006.
5. Pang-Ning Tan, Michael Steinbach and Vipin Kumar “Introduction to Data Mining”, Pearson Education, 2007.

DATA MINING LAB

List of Lab experiments:

1. Installation of WEKA Tool
2. Creating new Arff File
3. Pre-Processes Techniques on Data Set 16 Pre-process a given dataset based on Handling Missing Values
4. Generate Association Rules using the Apriori Algorithm
5. Generating association rules using fp growth algorithm
6. Build a Decision Tree by using J48 algorithm
7. Naïve bayes classification on a given data set
8. Applying k-means clustering on a given data set
9. Calculating Information gains measures
10. OLAP Cube and its different operations
11. Case Study: Create Student. ariff file to suggest better college using Decision tree
12. Case Study: Create Placement. ariff file to identify the students who are eligible for placements using KNN

ARTIFICIAL INTELLIGENCE

Semester	2			
Course code	CC-A06			
Category	Core Course			
Course title	Artificial Intelligence			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5			
Practical External	20			
Total	100 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. Learn the concept of Artificial intelligence, problem solving with example and searching process.
2. Understand basic concepts of Expert system with its architecture and development life cycle.
3. Understand the concepts of knowledge, acquisition of knowledge and various levels and schemes with the help of which knowledge can be represented.
4. Learn the concepts of perception, basic concepts of Neural network, learning in neural
5. Handle the uncertainty in knowledge using fuzzy logic and understand various concepts of fuzzy logic.

UNIT - I

Problem solving: State space search: Production systems, Search space control, Depth first search, unknown search, Hill climbing best first search, branch and bound. Best First Search, Problem Reduction, Constraints, Satisfaction, Means End Analysis.

UNIT - II

Knowledge Representation: Predicate logic: Skolemzing queries, Unification, Modus pones, Resolution, dependency directed back tracking.

Rule Based Systems: Forward reasoning Conflict resolution, Backward reasoning. Use of non back track

UNIT - III

Perception: Sensing, Speech recognition, Vision, Action, Neural networks: Introduction, Comparison of artificial neural networks with biological neural networks, Learning in neural networks, Perceptions, Back propagation networks, application of neural networks.

Fuzzy logic: Definition, Difference between Boolean and Fuzzy logic, fuzzy subset, fuzzy membership function, fuzzy expert system, Inference process for fuzzy expert system, fuzzy controller

UNIT - IV

Expert system development life cycle: Problem selection, Prototype construction, Formalization, Implementation, Evaluation, Knowledge acquisition: Knowledge engineer, Cognitive behavior, Acquisition techniques.

Knowledge representation: Level of representation, Knowledge representation schemes, Formal logic, Inference Engine, Semantic net, Frame, Scripts

TEXT AND REFERENCE BOOKS:

1. Rich Elaine and Knight Kevin : Artificial Intelligence, Tata McGraw Hill.
2. Tani Moto : Introduction to AI using LISP.
3. Patterson : Artificial Intelligence and Expert Systems.
4. Winston, P.H. and: LISP B.K.P.
5. Sangal Rajeev : LISP Programming, Tata McGraw Hill.
6. Balagurusamy : Artificial Intelligence & Technology.
7. Mishkoff, Henry C: Understanding Artificial Intelligence, BPB Publ.
8. Bharti & Chaitenya: Natural Language Processing, PHI

ARTIFICIAL INTELLIGENCE LAB

List of lab experiments

1. Write code to perform DFS and BFS on a given state space and compare their efficiencies.
2. Implement hill climbing algorithms for a given problem and analyze it.
3. Write a Program to find the solution for travelling salesman Problem
4. Apply problem reduction and means-end analysis to a complex problem and document the steps and results.
5. Implement a unification algorithm and apply it to logical statements to perform inference.
6. Develop a simple rule-based system and implement forward reasoning and backward reasoning mechanisms.
7. Use a neural network framework to build a simple recognition system.
8. Implement a neural network to classify images from a dataset like MNIST or CIFAR-10.
9. Design a fuzzy controller for a simple system, such as a temperature control system, and implement it.
10. Select a problem domain, acquire knowledge, and develop a prototype expert system.

INFORMATION RETRIEVAL SYSTEMS

Semester	2			
Course code	DSE-02			
Category	Discipline Specific Elective Course			
Course title	Information Retrieval Systems			
Scheme and Credits	L	T	P	Credits
	2	1	0	3
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	-			
Practical External	-			
Total	75 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. Understand basic Information Retrieval Systems and learn how Boolean queries are processed.
2. Realize the data structures like Inverted Indices used in Information retrieval systems.
3. understand the basic concept of Search Engines their architecture and its various functional components and understand the basic concept of Web crawlers and their architecture
4. identify the different types of indices: inverted index, positional index, biword index and be able make estimations and model distribution of terms and compressions
5. enumerate various types of indices and also understand the concept of efficient storage of indices and learn tf-idf scoring and vector space model scoring for ranking

UNIT - I

Information retrieval problem, an inverted index, Processing Boolean queries, The extended Boolean model versus ranked retrieval, an inverted index, Bi-word indexes, Positional indexes, Combination schemes.

UNIT – II

Search Engines: Basic Building Blocks and Architecture, Text Acquisition, Text Transformation, Index Creation, User Interaction, Ranking, Evaluation.

CRAWL AND FEEDS: Crawling the Web, Retrieving Web Pages, The Web Crawler, Freshness, Focused Crawling, Deep Web, Crawling Documents and Email, Storing the Documents, Detecting Duplicates

UNIT – III

INDEX CONSTRUCTION AND COMPRESSION: Hardware basics, Blocked sort-based indexing, Single-pass in-memory indexing, Distributed indexing, Dynamic indexing

Index compression: Statistical properties of terms in information retrieval, Heaps' law: Estimating the number of terms, Zipf's law: Modeling the distribution of terms, Dictionary compression, Dictionary as a string, Blocked storage, Postings file compression

UNIT - IV

SCORING, TERM WEIGHTING AND THE VECTOR SPACE MODEL: Parametric and zone indexes, Weighted zone scoring, Learning weights, The optimal weight, Term frequency and weighting, Inverse document frequency, Tf-idf weighting, The vector space model for scoring , Computing scores in a complete search system.

TEXT AND REFERENCE BOOKS:

1. C.D.Manning, P. Raghavan and H.Schutze ―Introduction to Information Retrieval‖, Cambridge University Press, Latest Edition
2. B.Croft, D.Metzler, T.Strohman, ―Search Engines : Information Retrieval in Practice‖,AddisionWesley, Latest Edition

Semester 3

DESIGN AND ANALYSIS OF ALGORITHMS

Semester	3			
Course code	CC-A07			
Category	Core Course			
Course title	Design and Analysis of Algorithms			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

***Note**: The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. Prove the correctness and analyze the running time of the basic algorithms for those classic problems in various domains.
2. Analyze worst-case running times of algorithms using asymptotic analysis
3. Explain the major graph algorithms and their analyses. Employ graphs to model engineering problems, when appropriate.
4. Compare between different data structures. Pick an appropriate data structure for a design situation..
5. Apply the algorithms and design techniques to solve problems.

UNIT-I

Sets and disjoint: sets, union, sorting and searching algorithms and their analysis in terms of space and time complexity.

Divide and Conquer: General method, binary search, merge sort, quick sort, selection sort, Strassen's matrix multiplication algorithms and analysis of algorithms for these problems.

UNIT-II

Greedy Method: General method, Knapsack problem, Job sequencing with deadlines, Minimum spanning trees- Prim's and Kruskal's algorithms, Single source paths- Dijkastra algorithms and analysis of these problems.

Dynamic Programming: General method, Optimal binary search trees, 0/1 Knapsack, Traveling Salesperson Problem.

UNIT-III

Back Tracking: General method, 8 Queen's Problem, Graph coloring, Hamiltonian cycles and analysis of these problems.

Branch and Bound: Method, 0/1 Knapsack and Traveling Salesperson Problem, efficiency considerations.

UNIT-IV

NP Hard and NP Complete Problems: Basic concepts, Cook's theorem, NP hard graph and NP scheduling problems some simplified NP hard problems.

Advanced data structures: Red-Black trees, B-trees, Fibonacci Heaps

TEXT BOOKS:

1. Fundamental of Computer algorithms, Ellis Horowitz and Sartaj Sahni, Galgotia Publ.
2. Introduction to Algorithms, Thomas H Cormen, Charles E Leiserson And Ronald L Rivest: TMH.

REFERENCE BOOKS:

1. The Design and Analysis of Computer Algorithm, Aho A.V. Hopcroft J.E., Addison Wesley.
2. Algorithms-The Construction, Proof and Analysis of Programs, Berlion, P.Bizard, P., Johan Wiley & Sons.
3. Writing Efficient Programs, Bentley, J.L., PHI.
4. Introduction to Design and Analysis of Algorithm, Goodman, S.E. & Hedetnieni, MGH.
5. Introduction to Computers Science- An algorithms approach , Jean Paul Trembley, Richard B.Bunt, T.M.H.
6. Fundamentals of Algorithms: The Art of Computer Programming Voll, Knuth, D.E., Naresh Publ.

DESIGN AND ANALYSIS OF ALGORITHMS LAB

List of Lab experiments

1. Write a program to implement different sorting techniques.
    - Bubble Sort
    - Insertion Sort
    - Selection Sort
    - Quick Sort
    - Merge Sort
2. Write a program to find minimum cost spanning tress.
3. Write a program to implement travelling sales person problem.
4. Write a program to find Longest Path in a Directed Acyclic Graph.
5. Write a program for Shortest path with exactly k edges in a directed and weighted graph.
6. Write a program find maximum number of edge disjoint paths between two vertices
7. Implement 0/1 Knapsack problem using Dynamic Programming.
8. Perform various tree traversal algorithms on tree of choice
9. Implement N-Queens Problem

MACHINE LEARNING -I

Semester	3			
Course code	CC-A08			
Category	Core Course			
Course title	Machine Learning -I			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

***Note**: The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit..*

COURSE OUTCOMES:

1. Analyze methods and theories in the field of machine learning and provide an introduction to the basic principles, techniques, and applications of machine learning, supervised, unsupervised and reinforcement learning.
2. Comprehend and apply regression techniques.
3. Comprehend and implement various classification and clustering methods.
4. Understand the concept of association rule mining and neural networks and their implementation in context of Machine Learning.

UNIT-I

Introduction: Well-Posed learning problems, Basic concepts, Designing a learning system, Issues in machine learning. Types of machine learning: Learning associations, Supervised learning, Unsupervised learning and Reinforcement learning.

Data Pre-processing: Need of Data Pre-processing, Data Pre-processing Methods: Data Cleaning, Data Integration, Data Transformation, Data Reduction; Feature Scaling (Normalization and Standardization), Splitting dataset into Training and Testing set.

UNIT-II

Regression: Linear Regression, Multiple Linear Regression and Polynomial Regression, Evaluating Regression Models‟ Performance (RMSE, Mean Absolute Error, Correlation, RSquare).

Classification: Need and Applications of Classification, Logistic Regression, Decision tree, Tree induction algorithm – split algorithm based on information theory, split algorithm based on Gini index; Random forest classification, Naïve Bayes algorithm; K-Nearest Neighbours (K-NN), Evaluating Classification Models‟ Performance (Sensitivity, Specificity, Precision, Recall, etc)

UNIT-III

Artificial Neural Network: Need and Application of Artificial Neural Network, Neural network representation and working, Activation Functions, Forward Propagation, Backward Propagation, Gradient decent.

UNIT-IV

Clustering: Need and Applications of Clustering, Partitioned methods, Hierarchical methods, Density-based methods.

Association Rules Learning: Need and Application of Association Rules Learning, Basic concepts of Association Rule Mining, Naïve algorithm, Apriori algorithm.

Text Book(s)

1. Ethem Alpaydin, Introduction to Machine Learning, MIT Press, Prentice Hall of India, Third Edition 2014.
2. Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning series) 2nd edition, A Bradford Book;2018.

Reference Books

1. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, Foundations of Machine Learning, MIT Press, 2012.
2. Tom Mitchell, Machine Learning, McGraw Hill, 3rd Edition, 1997.
3. Charu C. Aggarwal, Data Classification Algorithms and Applications, CRC Press, 2014

MACHINE LEARNING-I LAB

List of Experiments

3. Linear & Multiple Linear Regression
4. Naïve Bayes classifier
5. Decision trees – ID3 & CART
6. Logistic regression
7. Support Vector Machines – Linear & Non-linear
8. Single & Multilayer Perceptron
9. K-NN and K-Means clustering
10. Random forest implementation
11. Principal component analysis

FUZZY SYSTEMS AND APPLICATIONS

Semester	3			
Course code	CC-A09			
Category	Core Course			
Course title	Fuzzy Systems and Applications			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

At the end of this course, students will demonstrate the ability to

1. Interpret fuzzy set theory and uncertainty concepts.
2. Identify the similarities and differences between probability theory and fuzzy set theory and their application conditions.
3. Apply fuzzy set theory in modeling and analyzing uncertainty in a decision problem.
4. Apply fuzzy control by examining simple control problem example.

UNIT - I

Fuzzy logic: Definition, Difference between Boolean and Fuzzy logic, fuzzy subset, fuzzy membership function, Classical sets vs Fuzzy Sets - Need for fuzzy sets, Mathematical representations - Level Sets - Fuzzy functions - Zadeh’s Extension Principle.

UNIT - II

Operations on [0,1] - Fuzzy negation, triangular norms, t- conorms, fuzzy implications, Aggregation Operations, Fuzzy Functional Equations

Fuzzy Binary and n-ary relations - composition of fuzzy relations - Fuzzy Equivalence Relations - Fuzzy Compatibility Relations - Fuzzy Relational Equations,

UNIT - III

Fuzzy Measures - Evidence Theory - Necessity and Belief Measures - Probability Measures vs Possibility Measures, Fuzzy Decision Making - Fuzzy Relational Inference - Compositional Rule of Inference - Efficiency of Inference - Hierarchical

UNIT - IV

Fuzzy If-Then Rule Base - Inference Engine - Takagi-Sugeno Fuzzy Systems - Function Approximation Applications Advanced topics: Adaptive fuzzy inference systems: Adaptive networks - Architectures - Learning rules. Adaptive neuro-fuzzy inference systems (ANFIS) - Architectures - Hybrid learning rules.

Text Books:

1. George J Klir and Bo Yuan, “Fuzzy Sets and Fuzzy Logic : Theory and Applications”, Prentice Hall NJ,1995.
2. Timothy J. Ross, “Fuzzy Logic with Engineering Applications”, 3rd Edition, Willey, 2010.

References:

1. E P Klement, R Mesiar and E. Pap, Triangular norms, Kluwer Academic Press, Dordrecht, 2000.
2. H.J. Zimmermann, Fuzzy Set Theory and its Applications, Allied Publishers, New Delhi, 1991.
3. Kevin M Passino and Stephen Yurkovich, Fuzzy Control, Addison Wesley Longman, 1998.
4. M Grabisch et al., Aggregation Functions, Series - Encyclopedia Of Mathematics And Its Applications, Cambridge University Press, 2009
5. Michal Baczynski and Balasubramaniam Jayaram, Fuzzy Implications, Springer Verlag, Heidelberg, 2008.

FUZZY SYSTEMS AND APPLICATIONS LAB

List of lab experiments

1. Create membership functions (e.g., triangular, trapezoidal) for variables such as temperature, speed, and humidity. Compare and contrast with Boolean logic.
2. Implement both classical set operations and fuzzy set operations on sample datasets and analyze the results.
3. Implement fuzzy functions for operations such as addition and multiplication and observe their behavior with fuzzy inputs.
4. Perform operations on fuzzy sets using different T-Norms and T-Conorms and compare their effects.
5. Implement different fuzzy implication methods and aggregation operations on fuzzy sets.
6. Create fuzzy equivalence relations and compatibility relations. Implement and analyze the composition of fuzzy relations.
7. Implement and compare necessity and belief measures on given datasets.
8. Design a fuzzy decision-making system for a multi-criteria decision-making problem and implement compositional rule of inference.
9. Design a fuzzy rule-based system using Takagi-Sugeno fuzzy systems for function approximation applications.
10. Implement an ANFIS model for a given dataset, train the model using hybrid learning rules, and analyze its performance.

INTERNET OF THINGS

Semester	3			
Course code	DSE-03			
Category	Discipline Specific Elective Course			
Course title	Internet of Things			
Scheme and Credits	L	T	P	Credits
	2	1	0	3
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	-			
Practical External	-			
Total	75 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit..*

COURSE OUTCOMES:

1. Understand the concepts of Internet of Things
2. Analyze basic protocols network
3. Understand the concepts of Web of Things
4. Basic Understanding of Cloud Computing
5. Design IoT applications in different domain and be able to analyze their performance.

UNIT – 1

INTRODUCTION TO IOT: Introduction to IoT, Characteristics of IoT, Physical design of IoT, Logical design of IoT, Functional blocks of IoT, Communication models & APIs ,IoT & M2M Machine to Machine, Difference between IoT and M2M, Software define Network, Challenges in IoT(Design ,Development, Security)

UNIT – 2

NETWORK AND COMMUNICATION ASPECTS: Wireless medium access issues, MAC protocol survey, Survey routing protocols, Sensor deployment & Node discovery, Data aggregation & dissemination.

UNIT – 3

WEB OF THINGS: Web of Things vs Internet of things, two pillars of web, Architecture and standardization of IoT, unified Multitier-WoT architecture, WoT portals and Business intelligence, Cloud of things: Grid/SOA and cloud computing, Cloud middleware, cloud standards

UNIT – 4

RESOURCE MANAGEMENT IN IOT: Domain specific applications of IoT, Home automation, Industry applications, Surveillance applications, Other IoT applications Clustering, Synchronization, Software agents.

TEXT AND REFERENCE BOOKS:

1. Vijay Madisetti, Arshdeep Bahga, “Internet of Things: A Hands-On Approach”
2. Waltenegus Dargie, Christian Poellabauer, "Fundamentals of Wireless Sensor Networks: Theory and Practice"
3. Cuno Pfister, “Getting Started with the Internet of Things”, Shroff Publisher/Maker Media.
4. Internet of Things, RMD Sundaram Shriram K Vasudevan, Abhishek S Nagarajan, John Wiley and Sons

Semester 4

BIG DATA ANALYTICS

Semester	4			
Course code	CC-A10			
Category	Core Course			
Course title	Big Data Analytics			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. Explore the fundamental concepts of Big Data analysis
2. Identify and successfully apply appropriate techniques and tools to solve actual Big Data problems (derive value from vast data sets).
3. Examine the distributed and parallel computing and its application for big data analysis.
4. Analyse how to deal with huge amount of data and propose scalable solutions.
5. Evaluate statistical packages and deriving intelligence from unstructured information
6. Compile and contrast among different big data analytics tools and how they can help solving Industry challenges.

UNIT - I

Introduction to Big Data, challenges of conventional systems, Evolution of analytic scalability , Modern data analytic tools Modelling techniques: Mining frequent itemsets, Apriori algorithm, Handling large data sets in main memory ,Clustering techniques, clustering for parallelism, Classification and Prediction: Decision Tree induction, Developing models using Decision Tree Algorithms

UNIT - II

Frameworks: Overview of Hadoop, Hadoop Distributed File System, HDFS design and architecture Hadoop Map reduce Framework, HBASE, Interacting HDFS using HIVE, sample programs in HIVE-PIG

UNIT - III

Data Analysis and mining data streams: Regression modelling, Rule Induction Fuzzy decision trees and neural networks Introduction to streams concepts, Real time analytics platform, case studies

UNIT - IV

Visualization: Visual data analysis techniques, Interaction techniques Analytics using statistical packages, association intelligence from unstructured information Text analytics, industry challenges and application of analytic

TEXT BOOKS:

1. Bill Franks, “Taming the big data tidal wave: finding opportunities in huge data streams with advanced analytics”, John Wiley & Sons,2012
2. Anand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, Cambridge University Press,2012
3. Michael Berthold, David J. Hand, “Intelligent Data Analysis”, Springer 2007
4. Jiwaei Han, Micheline Kamber, “Data Mining Concepts and Techniques”, Second Edition, Elsevier, Reprinted 2008

BIG DATA ANALYTICS LAB

List of Subject related Experiments:

1. Install Apache Hadoop
2. Develop a MapReduce program to calculate the frequency of a given word in agiven file.
3. Develop a MapReduce program to find the maximum temperature in each year.
4. Develop a MapReduce program to find the grades of student's.
5. Develop a MapReduce program to implement Matrix Multiplication.
6. Develop a MapReduce to find the maximum electrical consumption in each year given electrical consumption for each month in each year.
7. Develop a MapReduce to analyze weather data set and print whether the day is shinny or cool day.

DEEP LEARNING

Semester	4			
Course code	CC-A11			
Category	Core Course			
Course title	Deep Learning			
Scheme and Credits	L	T	P	Credits
	3	0	2	4
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	5 marks			
Practical External	20 marks			
Total	100 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit.*

COURSE OUTCOMES:

1. Understand the fundamentals of deep learning and the main research activities in this field.
2. Emphasizing knowledge on various deep learning algorithms.
3. Understanding of CNN and RNN to model for real-world applications.
4. Understanding the various challenges involved in designing deep learning algorithms for varied applications.
5. Implement deep learning algorithms and solve real-world problems.
6. Identify the deep learning algorithms which are more appropriate for various types of learning tasks in various domains.

UNIT - I

Introduction: Definition, History of Deep Learning Deep Learning

Applications: Large-Scale Deep Learning, Computer Vision, Speech Recognition, Natural Language Processing

UNIT - II

Artificial Neural Networks: McCulloch-Pitts unit and Thresholding logic, Linear Perceptron, Perceptron Learning Algorithm, Linear separability. Convergence theorem for Perceptron Learning Algorithm. Feed forward Networks: Multilayer Perceptron, Gradient Descent, Back propagation, Empirical Risk Minimization, regularization.

Convolutional Networks: The Convolution Operation - Variants of the Basic Convolution Function - Structured Outputs - Data Types - Efficient Convolution Algorithms - Random or Unsupervised Features- LeNet, AlexNet

UNIT - III

Recurrent Neural Networks: Bidirectional RNNs - Deep Recurrent Networks Recursive Neural Networks - The Long Short-Term Memory and Other Gated RNNs

Generative Adversial Networks (GANs): Introduction, Discriminator, Generator, Activation, Common activation functions for GANs, BCE loss, Conditional GANs, Controllable generation, real life GANs

UNIT - IV

Deep Generative Models: Boltzmann Machines - Restricted Boltzmann Machines - Introduction to MCMC and Gibbs , Deep Belief Networks- Deep Boltzmann Machines

Optimization for Train Deep Models: Challenges in Neural Network Optimization, Basic Algorithms, Parameter Initialization Strategies, Algorithms with Adaptive Learning Rates, Approximate Second-Order Methods, Optimization Strategies and Meta-Algorithms.

TEXT AND REFERENCE BOOKS:

1. Nikhil Buduma, “Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithm”, O’Reilly, 2017.
2. Ian Goodfellow, YoshuaBengio and Aaron Courville, “Deep Learning”, MIT Press, 2016.
3. AurélienGéron, “Hands-On Machine Learning with Scikit- Learn and TensorFlow”, O’Reilly, 2017.
4. Nikhil Ketkar, “Deep Learning with Python: A Hands-on Introduction”, Apress, 2017.
5. Tariq Rashid, "Make your own neural network ", 2017.

DEEP LEARNING LAB

List of lab experiments

1. Setting up the Spyder IDE Environment and Executing a Python Program
2. Plotting of Activation Functions: Threshold functions, Signum function, Sigmoid function, Tanhyperbolic function, Ramp function, Identity function.
3. Implementation of linearly separable concept for a problem.
4. Implementation of some basic model like MCP with suitable example
5. Installing Keras, TensorFlow and Pytorch libraries and making use of them
6. Applying the Convolution Neural Network on computer vision problems
7. Image classification on MNIST dataset (CNN model with Fully connected layer)
8. Applying the Deep Learning Models in the field of Natural Language Processing.
9. Train a sentiment analysis model on IMDB dataset, use RNN layers with LSTM/GRU notes.
10. Applying the Autoencoder algorithms for encoding the real-world data
11. Applying Generative Adversial Networks for image generation and unsupervised tasks.

COMPUTATIONAL INTELLIGENCE

Semester	4			
Course code	DSE-04			
Category	Discipline Specific Elective Course			
Course title	Computational Intelligence			
Scheme and Credits	L	T	P	Credits
	2	1	0	3
Theory Internal	25 marks			
Theory External	50 marks			
Practical Internal	-			
Practical External	-			
Total	75 marks			
Duration of Exam	3 hrs			

Note: *The examiner will set nine questions in total. Question one will have seven parts from all units and the marks of first question will be of 20% of total marks of Question Paper and the remaining eight questions to be set by taking two questions from each unit and the marks of each question from Question no.2 to 9 will be of 20% of total marks of Question paper . The students have to attempt five questions in total, the first being compulsory and selecting one from each unit..*

COURSE Outcomes:

1. Provide a basic exposition to the goals and methods of Computational Intelligence
2. Study of the design of intelligent computational techniques.
3. Apply the Intelligent techniques for problem solving
4. Improve problem solving skills using the acquired knowledge in the areas of, reasoning, natural language understanding, computer vision, automatic programming and machine learning

UNIT - I

Introduction to Artificial Intelligence-Search-Heuristic Search-A* algorithm-Game Playing Alpha-Beta Pruning-Expert systems-Inference-Rules-Forward Chaining and Backward Chaining- Genetic Algorithms

UNIT - II

Proposition Logic - First Order Predicate Logic – Unification – Forward Chaining - Backward Chaining - Resolution – Knowledge Representation - Ontological Engineering - Categories and Objects – Events - Mental Events and Mental Objects – Reasoning Systems for Categories - Reasoning

with Default Information - Prolog Programming.

Non monotonic reasoning-Fuzzy Logic-Fuzzy rules-fuzzy inference-Temporal Logic Temporal Reasoning-Neural Networks-Neuro-fuzzy Inference

UNIT - III

Probability basics - Bayes Rule and its Applications - Bayesian Networks – Exact and Approximate Inference in Bayesian Networks - Hidden Markov Models - Forms of Learning - Supervised Learning - Learning Decision Trees – Regression and Classification with Linear Models - Artificial Neural Networks – Nonparametric Models - Support Vector Machines - Statistical Learning - Learning with Complete Data - Learning with Hidden Variables- The EM Algorithm Reinforcement Learning.

UNIT - IV

Natural language Processing-Morphological Analysis-Syntax Analysis-Semantic Analysis. AIl applications – Language Models - Information Retrieval – Information Extraction – Machine Translation – Machine Learning - Symbol-Based – Machine Learning: Machine Learning toolkits/Libraries.

TEXT BOOKS:

1. Stuart Russell, Peter Norvig, ―Artificial Intelligence: A Modern Approach‖, Third Edition, Pearson Education / Prentice Hall of India, 2010.
2. Elaine Rich and Kevin Knight, ―Artificial Intelligence‖, Third Edition, Tata McGrawHill, 2010.

REFERENCES BOOKS:

1. Patrick H. Winston. "Artificial Intelligence", Third edition, Pearson Edition, 2006.
2. Dan W.Patterson, ―Introduction to Artificial Intelligence and Expert Systems‖, PHI, 2006.
3. Nils J. Nilsson, ―Artificial Intelligence: A new Synthesis‖, Harcourt Asia Pvt. Ltd., 2000.