
# DEPARTMENT OF COMPUTER SCIENCE
# KAZI NAZRUL UNIVERSITY

P.S, Nazrul Road, Kalla Bypass More, North,
P.O, C H Kalla, Asansol, West Bengal 713340

## M.Sc in Computer Science (Data Science)

(Applicable form 2020-21 onwards)

**Course Name:** M.Sc in Computer Science (Data Science)

**Course Structure:**

Summary:

| Semester No | Total Credit |
|---|---|
| 1 | 22 |
| 2 | 24 |
| 3 | 20 |
| 4 | 22 |
| Total | 88 |

**Detailed Course Structure**

**Semester – I**

| Course Name | Course Type | Course Code | L - T - P | Course Credit | Sem Credit |
|---|---|---|---|---|---|
| Mathematics for Data Science | C | MSCCSC101 | 4 - 0 - 0 | 4 | 22 |
| Object Oriented Analysis and Design for Data Analytics | C | MSCCSC102 | 4 - 0 - 0 | 4 | |
| Advanced Database Management Systems | C | MSCCSC103 | 4 - 0 - 0 | 4 | |
| Advanced Operating Systems | C | MSCCSC104 | 4 - 0 - 0 | 4 | |
| Analytics Lab | C | MSCCSC105 | 0 - 0 - 4 | 2 | |
| DBMS Lab | C | MSCCSC106 | 0 - 0 - 4 | 2 | |
| OS Lab using Linux/Unix | C | MSCCSC107 | 0 - 0 - 4 | 2 | |

**Semester – II**

| Course Name | Course Type | Course Code | L - T - P | Course Credit | Sem Credit |
|---|---|---|---|---|---|
| Advanced Data Structures and Algorithms | C | MSCCSC201 | 4 - 0 - 0 | 4 | 24 |
| Computer Networks | C | MSCCSC202 | 4 - 0 - 0 | 4 | |
| Artificial Intelligence | C | MSCCSC203 | 4 - 0 - 0 | 4 | |
| Data Mining and Data Warehouse | C | MSCCSC204 | 4 - 0 - 0 | 4 | |
| AI Lab | C | MSCCSC205 | 0 - 0 - 4 | 2 | |
| Data Mining Lab | C | MSCCSC206 | 0 - 0 - 4 | 2 | |
| Minor Elective | MIE | See Pool | 4 - 0 - 0 | 4 | |

**Semester – III**

| Course Name | Course Type | Course Code | L - T - P | Course Credit | Sem Credit |
|---|---|---|---|---|---|
| Big Data Technology | C | MSCCSC301 | 4 - 0 - 0 | 4 | 20 |
| Machine Learning | C | MSCCSC302 | 4 - 0 - 0 | 4 | |
| Advanced Statistical Methods | C | MSCCSC303 | 4 - 0 - 0 | 4 | |
| Big Data Technology and OLTP Lab | C | MSCCSC304 | 0 - 0 - 4 | 2 | |
| Machine Learning and Advanced Analytics Lab | C | MSCCSC305 | 0 - 0 - 4 | 2 | |
| Minor Elective | MIE | See Pool | 4 - 0 - 0 | 4 | |

**Semester – IV**

| Course Name | Course Type | Course Code | L - T - P | Course Credit | Sem Credit |
|---|---|---|---|---|---|
| Data Security and Privacy | C | MSCCSC401 | 4 - 0 - 0 | 4 | 22 |
| Internet of Things | MJE (Any Three) | MSCCSMJE 401 | 4 - 0 - 0 | 4+4+4 =12 | |
| Green Computing | | MSCCSMJE 402 | 4 - 0 - 0 | | |
| Cloud Computing | | MSCCSMJE 403 | 4 - 0 - 0 | | |
| Computer Vision | | MSCCSMJE 404 | 4 - 0 - 0 | | |
| Deep Learning | | MSCCSMJE 405 | 4 - 0 - 0 | | |
| Natural Language Processing | | MSCCSMJE 406 | 4 - 0 - 0 | | |
| Project & Grand Viva | C | MSCCS402 | 0 - 0 - 12 | 6 | |

**Minor elective courses**

| Semester | Course Name | Course Code | L - T - P | Course Credit |
|---|---|---|---|---|
| II | Fundamentals of Programming (C) | MSCCSMIE201 | 4 - 0 - 0 | 4 |
| III | Business Intelligence | MSCCSMIE301 | 4 - 0 - 0 | 4 |

**Semester - I**

**Course Name: Mathematics for Data Science**

**Course Code: MSCCSC101**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

1. Learn different types of matrices, determinant, concept of rank, method of matrix inversion and their applications
2. Understand the concept of eigen values and eigen vectors
3. Understand the basic concept of LPP and applications of LPP
4. Learn about various measures of central tendency, dispersion, moments, skewness and kurtosis.
5. Understand simple and rank correlation, fitting of linear and quadratic regressions using principle of least squares

## Module 1: Basics of Linear Algebra

Matrices, Types of matrix, matrix operations and their properties, Symmetric, Skew-symmetric and Orthogonal matrices and their properties, Determinants and their properties, minors and co-factors, adjoint and reciprocal determinants, Jacobi's theorem, inverse and rank of a matrix, solution of System of linear equations , matrix method, Cramer's Rule, consistency and inconsistency of system of linear equations, Eigen values, Eigen vectors, Cayley-Hamilton Theorem , similar matrices, Diagonalization of a matrix, Solutions of system of linear equations using Gaussian eliminations and LU-decompositions.

## Module 2: Basics of LPP and Applications

Introduction, Various Components of LP Problem , General structure of LPP, Model Formulation of LPP, Basic and non-basic Variables, basic solution, Feasible Solution,Basic Feasible Solution, Degenerate and Non-degenerate Solution , Graphical Method of solution of LPP, Characteristics of an optimal solution: convex combination and Convex set and few important theorems, extreme points, convex hull and convex polyhedron.

## Module 2: Basics of Statistics

Measures of Central Tendency: mathematical and positional, Measures of Dispersion: range, quartile deviation, mean deviation, standard deviation, coefficient of variation, moments, measures of skewness and kurtosis.Bivariate distribution, covariance and correlation, rank correlation; fitting of linear and quadratic regression using principle of least squares.

**Text Books:**
1. Outline of statistics 1 and 2 by Goon-Gupta-Dasgupta
2. Fundamentals of mathematical statistics by Gupta and Kapoor
3. Complete Business Statistics Book By Amir Aczel

**Course Name: Object Oriented Analysis and Design for Data Analytics**

**Course Code: MSCCSC102**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

1. Understand theimportance and basic concepts of object oriented modeling,
2. Specify, analyze and design the use case driven requirements for a particular system.
3. Model the event driven states of objects and transform them into implementation specific layouts.
4. Identify, Analyze the subsystems, various components and collaborate them interchangeably.
5. Implement the object oriented model using an object oriented language

## Module 1: Object Oriented Analysis and Design

An Overview of Object Oriented Systems Development, Object Oriented Systems Development Life Cycle. Object Oriented methodologies, Rumbaugh Methodology - Booch Methodology - Jacobson Methodology, UML, Object Oriented Analysis & Design, software Quality and Usability, Case Studies

## Module 2: Object Oriented Language

Object Oriented Languages, Java and its features, Inheritance, exception handling, multithreading, Input/Output, Applet, Event Handling, Swing components, concept of JDBC.

**Text Books:**
1. Object-Oriented Analysis and Design by Sarnath Ramnath, Brahma Dathan, Springer
2. Object-Oriented Analysis And Design With Applications, 3/E by Booch
3. Java: The Complete Reference 7/E by Herbert Schildt, TMH

**Course Name: Advanced Database Management Systems**

**Course Code: MSCCSC103**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

1. Study the physical and logical database designs, database modeling, relational, hierarchical, and network models
2. Understand the different issues involved in the design and implementation of a database system.
3. Understand and use data manipulation language to query, update, and manage a database
4. Develop an understanding of essential DBMS concepts such as: database security, integrity, concurrency,
5. Design and build a simple database system and demonstrate competence with the fundamental tasks involved with modeling, designing, and implementing a DBMS.

## Module 1:

Overview of Database Management, Conceptual Database Design, Logical Database Design, Physical Database Design. Introduction to Relational Database : Relation, Optimization, The Catalog, Base Relvars and Views, Transactions, The Suppliers and Parts Database. Relational Model Concepts, Relational Model, Constraining, Referential Integrity Constraints, Defining Referential Integrity Constraints, Update Operations on Relations, Structured Query Language (SQL), Data Definition Language Commands, Data Manipulation Language Commands, Transaction Control Commands, SQL Command Syntax and Usage, The Basic Query Block, Querying Data with Multiple Conditions,
Basic Relational Algebra Operations, The Select Operation, Additional Relational Operations.

## Module 2:

ER- and EER-to-Relational Mapping: ER- to Relational Mapping Algorithm, Summary of Mapping for Model Constructs and Constraints Mapping EER Model Concepts to Relations, Query, Processing and Optimization: Query Processing, Query Optimization, Database Tuning.

## Module 3:

Object Oriented Database Systems: Characteristics of an Object-relation Database Management System (ORDBMS), Complex Objects, Inheritance, Function Overloading, Rules. Distributed Database : Distributed Database System, Distributed Database Design, Data Fragmentation, Data Replication, Data Allocation, Query Processing in Distributed Databases. Recovery : Transactions, Transaction Recovery, System Recovery, Media Recovery, Two-phase Commit. Database Models, Introduction to HDFS, Graph based Modeling,

**Text Books:**

1. Database System Concepts – 6th Edition by Silberschatz, Korth and Sudarshan
2. Fundamentals of Database Systems – 5th Edition by R.Elmasri, S. Navathe
3. Database Design and Relational Theory: Normal Forms and All That Jazz by C.J.

**Course Name: Advanced Operating System**

**Course Code: MSCCSC104**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

1. To understand the main components of an OS & their functions.
2. To study the process management and scheduling.
3. To understand various issues in Inter Process Communication (IPC) and the role of OS in IPC.
4. To understand the concepts and implementation Memory management policies and virtual memory. To understand the working of an OS as a resource manager, file system manager, process manager, memory manager and I/O manager and methods used to implement the different parts of OS
5. To study the need for special purpose operating system with the advent of new emerging technologies

## Module 1:

OS services and components, multitasking, multiprogramming, time sharing, buffering, Spooling Process & thread management, context switching, multithreading Concurrency control, mutual exclusion requirements, semaphores, monitors, Dead locks - detection, recovery, avoidance and prevention

## Module 2:

Memory management, partitioning, swapping, paging, segmentation, virtual memory, Demand paging, page replacement and allocation algorithm

## Module 3:

Introduction to Distributed Systems, Architectures of Distributed Systems, communication networks, Mutual Exclusion in Distributed Systems, RMI, concept of Replication, Distributed File Systems (NFS, AFS, coda) overview, security in Distributed Systems. HDFS File and Storage Management.

**Text Books:**

1. Distributed Operating systems by Andrew s.Tanenbanm
2. Operating System Concepts, 5th ed. by Silberschatz and Galvin
3. Advanced Concepts In Operating Systems by Mukesh Singhal and Niranjan Shivaratri

**Course Name: Analytics Lab**

**Course Code: MSCCSC105**

| Course Type: **Practical** | Course Details: Core | | L-T-P: **0-0-4** |
|---|---|---|---|
| Credit: 2 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Practical | Practical |
| | | **30** | **20** |

***Course Learning Outcomes:***

1. Getthe brief idea of what Python is and touch on the basics.
2. Learn data handling using Python
3. learn how to perform simple statistical analysis, create meaningful data visualizations, predict future trends from data, and more

## Module 1:

Python Installation

## Module 2:

Basic Data type, functions for reading and writing data, control structures, functions, loops, str function

## Module 3:

Debugging tools, simulation and profiling,

## Module 4:

Descriptive statistics and graphics, probability and distribution.

**Course Name: DBMS Lab**

**Course Code: MSCCSC106**

| Course Type: **Practical** | Course Details: Core | | L-T-P: 0**-0-4** |
|---|---|---|---|
| Credit: 2 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Practical | Practical |
| | | **30** | **20** |

***Course Learning Outcomes:***

1. Understand, appreciate and effectively explain the underlying concepts of database technologies
2. Design and implement a database schema for a given problem-domain
3. Normalize a database
4. Populate and query a database using SQL DML/DDL commands.
5. Declare and enforce integrity constraints on a database using a state-of-the-art RDBMS

## Module 1:

Introduction to SQL constructs. Review of Basic SQL statements Select, Project, Join, Describing Oracle tables , Restricting row returns Creating basic reports, Using the set commands, Adding prompts to queries

## Module 2:

Joining Oracle tables -Equi-join, Outer join Hiding joins by creating views,Using IN, NOT IN, EXISTS and NOTEXISTS, Subqueries, Exercise – write a subquery, Correlated subquery, Non-correlated subqueries

## Module 2:

Advanced SQL operators -Between operator , IN and NOT IN operators, Sub-queries-EXISTS clause, Using wildcards in queries (LIKE operator),Aggregation in SQL -Count(*),Sum, Avg, Min and max. Using the group by clause, SQL access methods, Review of Basic joining methods-Merge join, Hash Join, Nested Loop join..

**Course Name: OS Lab using Linux / Unix**

**Course Code: MSCCSC107**

<table>
<tr><td>Course Type: Practical</td><td colspan="2">Course Details: Core</td><td>L-T-P: 0-0-4</td></tr>
<tr><td rowspan="3">Credit: 2</td><td rowspan="3">Full Marks: 50</td><td>CA Marks</td><td>ESE Marks</td></tr>
<tr><td>Practical</td><td>Practical</td></tr>
<tr><td>30</td><td>20</td></tr>
</table>

***Course Learning Outcomes:***

1. Apply basic operations in shell scripts which are required for different applications.
2. Identify and understand concept of file systems in shell script
3. Apply concept of creating new process from parent process.
4. Apply concept of virtual file and execute basic commands on it
5. Design communication mechanisms IPC and pipe on linux

## Module 1:

Shell Programming-creating a script, making a script executable, shell syntax (Variables, conditions, control structures, functions, commands).

## Module 2:

Process-starting a process, conditions, control structures, functions, commands), waiting for a process, zombie process Semaphore-programming with Semaphore

**Semester – II**

**Course Name: Advanced Data Structures and Algorithms**

**Course Code: MSCCSC201**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

Students will be able to

1. define basic static and dynamic data structures and relevant standard algorithms for them: stack, queue, dynamically linked lists, trees, graphs, heap, priority queue, hash tables, sorting algorithms, min-max algorithm,
2. demonstrate advantages and disadvantages of specific algorithms and data structures,
3. select basic data structures and algorithms for autonomous realization of simple programs or program parts
4. determine and demonstrate bugs in program, recognise needed basic operations with data structures
5. evaluate algorithms and data structures in terms of time and memory complexity of basic operations.

## Module 1:

Algorithms: properties of good algorithms, Efficiency of algorithms, Time and Space complexity. Non-linear data structures: trees, Binary search tree- algorithms on BST, balanced treesAVL rotations, multi-way search trees- B Tree, B+ tree. Basic concepts of Red-Black tree, splay tree, tries. Elementary Graph algorithms- graph representations, Depth-first search, Breadth-first search, spanning tree.

## Module 2:

Algorithm design techniques-Divide and conquer strategy- general method, binary search, minmax algorithm, quicksort, Strassen‟s matrix multiplication. Greedy Method –general 3 method, knapsack problem, minimum cost spanning trees- Prim‟s algorithm, Kruskals algorithm, Single source shortest path.

## Module 3:

Dynamic Programming –all pairs shortest path, matrix chain multiplication, longest common subsequence. Backtracking- n-queens problem & solution. Branch and Bound – Travelling salesperson problem. Randomized algorithms- Randomized quicksort, Deterministic and non deterministic algorithms, NP hard and NP complete problems.

**Text Books:**


1. Horowitz, Sahni, Rajasekaran; Computer algorithms - Galgotia, 1998
2. Cormen, Thomas H; Leiserson, Charles E; Rivest, Ronald L, Introduction to Algorithms. - Prentice Hall of India, 1990.
3. Samanta D., Classic Data Structures, Prentice Hall of India.
4. Levitin, Introduction to the Design and Analysis of Algorithms
5. G.L. Heileman, Data Structutes, Algorithms and Object Oriented Programming
6. Horwitz ,E and Sahni, Sartaj, Fundamentals of Data structures.- Galgotia


**Course Name:Computer Networks**

**Course Code: MSCCSC202**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

Students will be able to

1. Independently understand basic computer network technology.
2. Understand and explain Data Communications System and its components.
3. Identify the different types of network topologies and protocols.
4. Enumerate the layers of the OSI model and TCP/IP. Explain the function(s) of each layer.
5. Identify the different types of network devices and their functions within a network
6. Understand and building the skills of subnetting and routing mechanisms.
7. Familiarity with the basic protocols of computer networks

## Module 1:

Introduction to Data Communication: Components of Data Communication, Networks, Protocols and Standards, Network Models, Review of Reference Models: OSI, TCP/IP and their comparison. Data and signals, Channel Capacity, Inter connecting devices – Repeater, Hub, Switch, Bridge, Router, Gateway.

## Module 2:

Mobile Communication - Introduction: Mobile and Wireless Devices, Simplified Reference Model, Need for Mobile Computing, Wireless Transmission, Multiplexing, Modulation, Spread Spectrum, Cellular Systems, Medium Access Control, Comparisons, Telecommunications System: GSM, DECT, TETRA, UMTS and IMT 2000, Satellite System, Broadcast Systems, Wireless LAN: IEEE 802.11, Hyper LAN, Bluetooth.

## Module 3:

Mobile Network and Transport Layers:- Mobile IP- Goals, Requirements, IP packet delivery, Advertisement and discovery. Registration, Tunneling and Encapsulation, Optimization, Reverse Tunneling, IPv6, Dynamic Host configuring protocol, Ad hoc networks - Routing, DSDV, Dynamic source routing. Traditional TCP, Classical TCP Improvements Indirect TCP, Wireless Application Protocol, World Wide Web. Overview of Wireless sensor networks, VPN, IOT and LiFi.

**Text Books:**


1. Behrouz A Forouzan, Data Communications and Networking, McGraw- Hill, 2006
2. Jochen Schiller, "Mobile Communication", Pearson Education, Delhi, 2000.
3. Brijendra Singh, Data Communication and Computer Networks, PHI, 2011.
4. Tanenbaum Andrew S., Computer Networks, TMH. • Kurose, James F., Ross Keith W., Computer Networking: A top-down approach featuring the internet, Addison-Wesley.
5. Comer, Computer Networks and Internet with Internet Applications, PHI, 2009.
6. Stallings William, Data and Computer Communication, Pearson, 2007.


### Course Name:Artificial Intelligence

**Course Code: MSCCSC203**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

The student shall be able to:

1) Demonstrate fundamental understanding of the history of artificial intelligence (AI) and its foundations.
2) Apply basic principles of AI in solutions that require problem solving, inference, perception, knowledge representation, and learning.
3) Demonstrate awareness and a fundamental understanding of various applications of AI techniques in intelligent agents, expert systems, artificial neural networks and other machine learning models.
4) Demonstrate proficiency developing applications in an 'AI language', expert system shell, or data mining tool.
5) Demonstrate proficiency in applying scientific method to models of machine learning.
6) Demonstrate an ability to share in discussions of AI, its current scope and limitations, and societal implications.

## Module 1:

Overview of Artificial Intelligence,Search and Control Strategies, General Problem solving, Production systems, Controlstrategies: forward and backward chaining Exhaustive searches: Depth first Breadthfirst search, Heuristic search techniques: Hill climbing, Branch and Bound

technique,Best first search and A* algorithm, AND/OR Graphs, Problem reduction and AO*algorithm, Constraint Satisfaction problems Game Playing Minmax search procedure,Alpha-Beta cutoffs,

## Module 2:

Knowledge Representation - Formalized Symbolic Logics. Dealing with Inconsistenciesand Uncertainties. Probabilistic Reasoning. Structured Knowledge: Graphs, Frames, andRelated Structures. Object-Oriented Representations.Additional Refinements, Matching Techniques. KnowledgeOrganization and ManagementBasic understanding of Fuzzy Logic,

## Module 3:

Artificial Neural Network, Perceptron, NaturalLanguage Processing, Pattern Recognition, overview on Expert Systems, AI Programming Languages – LISP/PROLOG

**Text Books:**


1. Introduction to Artificial Intelligence and Expert Systems by D.W. Patterson
2. Artificial Intelligence: A Modern Approach - 3rd edition by Stuart Russell & PeterNorvig
3. Artificial intelligence by Elaine Rich & Kevin Knight
4. Principles of Artificial Intelligence by J. Nilsson, Narosa Publishing House


### Course Name:Data Mining and Data Warehouse

**Course Code: MSCCSC204**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

Students will be able to:

1. Understand Data Warehouse fundamentals, Data Mining Principles
2. Design data warehouse with dimensional modelling and apply OLAP operations.
3. Identify appropriate data mining algorithms to solve real world problems
4. Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining
5. Describe complex data types with respect to spatial and web mining.
6. Benefit the user experiences towards research and innovation. integration.

## Module -1

Data Warehouse: Introduction to Data Ware House, Differences between operational data base systems and data Ware House, Data Ware House characteristics, Data Ware House Architecture and its components, Extraction-Transformation-Loading, Logical (MulitDimensional), Data Modeling, Schema Design, star and snow-Flake Schema, Fact Constellation, Fact Table, Fully Addictive, Semi-Addictive, Non-Addictive Measures; FactLess-Facts, Dimension Table characteristics; Fact-Less-Facts, Dimension Table

characteristics; OLAP cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.

## Module -2

Introduction to Data Mining: Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing- Data Cleaning, Missing Data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binaryzation, Data Transformation; Measures of similarity and dissimilarity-Basics.

## Module -3

Association Rules: Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation, APRIORI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent Item Set-Maximal Frequent Item Set, Closed Frequent Item Set.

## Module -4

Classification: Problem definition, General Approaches to solving a classification problem, Evaluation of Classifiers, Classification techniques, Decision trees-Decision Tree Construction, Methods for expressing attribute test conditions, Measures for Selecting the Best split, Algorithm for Decision tree Induction, Naïve-Bayes Classifier, Bayesian Belief Networks; K-nearest neighbor classification-Algorithm and characteristics.

## Module -5

Clustering: Problem Definition, Clustering overview, Evaluation of clustering algorithms, Partitioning clustering K-Means Algorithm, K-Means Additional Issues, PAM Algorithm, Hierarchical Clustering-Algorithm- Agglomerative Methods and Divisive Methods, Basic Agglomerative Hierarchical Clustering Algorithm, Specific techniques, Key Issues in Hierarchical Clustering, Strengths and weakness, Outlier Detection

**TEXT BOOK:**

1) Data Mining-Concepts and Techniques- Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier, 2 Edition, 2006.
2) Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbanch, Pearson Education.
3) Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
4) Data Ware Housing Fundamentals, Pualraj Ponnaiah, Wiley Student Edition.
5) The Data Ware House Life Cycle Toolkit- Ralph Kimball, Wiley Student Edition.
6) Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University

### Course Name: AI Lab

### Course Code: MSCCSC205

| Course Type: **Practical** | Course Details: Core | | L-T-P: **0-0-4** |
|---|---|---|---|
| Credit: 2 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Practical | Practical |
| | | **30** | **20** |

***Course Learning Outcomes:***

Students will be able to

1. Apply various AI search algorithms (uninformed, informed, heuristic, constraint satisfaction,)

7. Understand the fundamentals of knowledge representation, inference and theorem proving using AI tools
2. Demonstrate working knowledge of reasoning in the presence of incomplete and/or uncertain information
3. Ability to apply knowledge representation, reasoning, and machine learning techniques to real-world problems

## Module 1

Programming Language: PROLOG

1. Study of Prolog.
2. Write simple fact for the statements using PROLOG.
3. Write predicates One converts centigrade temperatures to Fahrenheit, the other checks if a temperature is below freezing.
4. Write a program to solve the Monkey Banana problem.
5. WAP in turbo prolog for medical diagnosis and show the advantage and disadvantage of green and red cuts.
6. WAP to implement factorial, fibonacci of a given number.
7. Write a program to solve 4-Queen problem.
8. Write a program to solve traveling salesman problem.
9. Solve any problem using depth first search.
10. Solve any problem using best first search.
11. Solve 8-puzzle problem using best first search

## Module 2

Programming Language: LISP

1. Study of LISP
2. Write a program to solve water jug problem using LISP
3. Do practise of more programs

**Course Name: Data Mining Lab**

**Course Code: MSCCSC206**

| Course Type: **Practical** | Course Details: Core | | L-T-P: **0-0-4** |
|---|---|---|---|
| Credit: 2 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Practical | Practical |
| | | **30** | **20** |

***Course Learning Outcomes:***

Students will be able to

1. The data mining process and important issues around data cleaning, pre-processing andintegration.
2. The principal algorithms and techniques used in data mining, such as clustering, associationmining, classification and prediction.

## Module 1

Manipulating strings, Processing Files, Manipulating Lists, Lists and Strings, Dictionaries, Counting with Dictionaries, Dictionaries and Files, Tuples, Tuples and Sorting, Regular Expressions,

## Module 2

Networked programs, Sockets and Applications, parsing HTML with Beautiful soup, parsing XML by python, REST, JSON and APIs, extracting data from JSON, Using database by python,

## Module 3

Object oriented python, Geocoding, Page rank and web searching, Gmane.

**Semester – III**

**Course Name: Big Data Technology**

**Course Code: MSCCSC301**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

1. Ability to identify the characteristics of datasets and compare the trivial data and big data for various applications.
2. To study the basic technologies that forms the foundations of Big Data.
3. To study the programming aspects of cloud computing with a view to rapid prototyping of complex applications.
4. To understand the specialized aspects of big data including big data application, and big data analytics.
5. To study different types Case studies on the current research and applications of the Hadoop and big data in industry

## Module 1:

Big data definition, structured and unstructured data.
Need for analytics, Big data programming ( Hadoop, Map-Reduce),

## Module 2:

Application Data store (NoSQL), OLAP.

## Module 3:

Optimization Techniques, Data flow framework.
Programming Map-Reduce, Best practices.

**Text Books:**

1. Handbook of big data technology by Zomaya and Sakr.

2. Real time Big Data Analytics Book by Sumit Gupta
3. Tom White " Hadoop: The Definitive Guide" Third Edit on, O'reily Media, 2012.
4. Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.

**Course Name: Machine Learning**

**Course Code: MSCCSC302**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***
students should be able to:

1. Develop an appreciation for what is involved in Learning models from data
2. Understand a wide variety of learning algorithms
3. Understand how to evaluate models generated from data
4. Apply the algorithms to a real problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models

## Module 1:

Introduction: Basic definitions, types of learning, hypothesis space and inductive bias, evaluation, cross-validation

## Module 2:

Linear regression, Decision trees, overfitting

## Module 3:

Instance based learning, Feature reduction, Collaborative filtering based recommendation Probability and Bayes learning

## Module 4:

Logistic Regression, Support Vector Machine, Kernel function and Kernel SVM

## Module 5:

Neural network: Perceptron, multilayer network, backpropagation, introduction to deep neural network

## Module 6:

Computational learning theory, PAC learning model, Sample complexity, VC Dimension, Ensemble learning

## Module 7:

Clustering: k-means, adaptive hierarchical clustering, Gaussian mixture model

**Text Books:**


1. Machine Learning. Tom Mitchell. First Edition, McGraw- Hill, 1997.
2. Introduction to Machine Learning Edition 2, by Ethem Alpaydin
3. Machine Learning and Knowledge Discovery edited by Walter Daelemans, Katharina Morik
4. Pattern Recognition and Machine Learning by Christopher Bishop
5. Introduction to Machine learning with python by Andreas C. Müller and Sarah Guido


**Course Name: Advanced Statistical Methods**

**Course Code: MSCCSC303**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

students will have the knowledge and skills to:

1. Describe the rationale behind the formulation and components of a statistical model.
2. Compare and contrast statistical models in the context of a particular scientific question.
3. Communicate statistical ideas to a diverse audience.
4. Formulate a statistical solution to real-data research problems.
5. Demonstrate an understanding of the theoretical and computational underpinnings of various statistical procedures, including common classes of statistical models.

## Module 1:

Recapitulation of basic statistical concepts,
Descriptive statistics (measures of location and dispersion),
Distributions
Graphical representation of data,
Contingency tables,

## Module 2:

Correlation and linear regression,
Hypothesis testing,
Fundamentals of modelling,
Multiple testing and the corresponding correction methods,
Graphical presentation of higher dimensional data,

## Module 1:

Multivariate regression, linear and polynomial

Analysis of variance including interaction,
Factor analysis,
Topics according to request

**Text Books:**


1. Moore, David S. (1991), Statistics: concepts and controversies, 3rd ed., New York: W.H. Freeman and Company
2. Muenchen, Robert A. (2011), R for SAS and SPSS Users, 2nd ed., New York et al.: Springer
3. Qian, Song S. (2010), Environmental an ecological statistics with R, New York: Taylor & Francis Group
4. Ross, Sheldon M. (2004), Introduction to probability and statistics for engineers and scientists, 3rd ed., Amsterdam et al.: Elsevier Academic Press


**Course Name: Big Data Technology and OLTP Lab**

**Course Code: MSCCSC304**

| Course Type: **Practical** | Course Details: Core | | L-T-P: **0-0-4** |
|---|---|---|---|
| Credit: 2 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Practical | Practical |
| | | **30** | **20** |

***Course Learning Outcomes:***

1. To study the basic technologies that forms the foundations of Big Data.
2. To understand the specialized aspects of big data including big data application, and big data analytics.
3. To study different types Case studies on the current research and applications of the Hadoop and big data in industry
4. To study Data warehousing and OLAP using some real life examples

## Module 1

A. NoSQL Lab using (MongoDB/Redis/Cassandra/CouchDB/Hbase using HDFs etc):

Introduction to Nosql,Difference between RDBMS to NOSQL,JSON and BSON documents,Introduction to MongoDB/.. and its Features,Database, Collection and Documents,Various Data Types in MongoDB/..,Introduction to mongo/.. shell,CRUD Operations,Database Operations,Read and Write Operations,Aggregation,Data Modeling Introduction,Data Modeling Concept,Storage Engine,Indexing,Replication Concept,Failover & Recovery

## Module 2

B. Multidimensional Data Modeling using OLAP:

Introduction of Datawarehousing and OLAP, example of a DataWarehouse and Data mart, Data Cleaning and integration, Data analysis techniques, Transformation algorithms, Integrations.

**Course Name: Machine Learning and Advanced Analytics Lab**

**Course Code: MSCCSC305**

| Course Type: **Practical** | Course Details: Core | | L-T-P: **0-0-4** |
|---|---|---|---|
| Credit: 2 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Practical | Practical |
| | | **30** | **20** |

***Course Learning Outcomes:***
Students will be able:

1. To introduce basic machine learning techniques.
2. To develop the skills in using recent machine learning software for solving practical problems in high-performance computing environment.
3. To develop the skills in applying appropriate supervised, semi-supervised orunsupervised learning algorithms for solving practical problems

## Module 1

1. Exercises to solve the real-world problems using the following machine learning methods:
    - Linear Regression
    - Logistic Regression
    - Multi-Class Classification
    - Neural Networks
    - Support Vector Machines
    - K-Means Clustering & PCA

2. Develop programs to implement Anomaly Detection & Recommendation Systems.
3. Implement GPU computing models to solving some of the problems mentioned in Problem 1.

**Semester – IV**

**Course Name: Data Security and Privacy**

**Course Code: MSCCSC401**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

1. Protect and defend computer systems and networks from cybersecurity attacks.
2. Diagnose and investigate cybersecurity events or crimes related to computer systems and digital evidence.
3. Effectively communicate in a professional setting to address information security issues.

## Module 1:

Computer Security Concepts, The OSI Security Architecture, Security Attacks, Security Services, Security Mechanisms, A Model for Network Security, Classical Encryption Techniques, Symmetric Cipher Model, Substitution Techniques, Transposition Techniques, Rotor Machines, Steganography, Cryptographic Tools, Confidentiality with Symmetric Encryption, Message Authentication and Hash Functions, Public-Key Encryption, Digital Signatures and Key Management, Random and Pseudorandom Numbers,

## Module 2:

Practical Application: Encryption of Stored Data, User Authentication, Means of Authentication, Password-Based Authentication, Token-Based Authentication, Biometric Authentication, Remote User Authentication, Security Issues for User Authentication, Malicious Software, Types of Malicious Software (Malware), Propagation—Infected Content—Viruses, Propagation—Vulnerability Exploit—Worms, Propagation—Social Engineering—SPAM Email, Trojans, Payload—System Corruption, Payload—Attack Agent—Zombie, Bots, Payload—Information Theft—Key loggers, Phishing, Spyware, Payload—Stealthing— Backdoors, Rootkits, Countermeasures,

## Module 3:

Firewalls and Intrusion Prevention Systems, the Need for Firewalls, Firewall Characteristic, Types of Firewalls, Firewall Basing, Firewall Location and Configurations, Intrusion Prevention Systems.

**Text Books:**

1. Cryptography and Network Security: Principles and Practice by William Stalings 6th Edition published by PHI (2011)

2. Computer security principles and practice, William Stallings, Lawrie Brown, third edition, Prentice-Hall, 2011

**Course Name: Internet of Things**

**Course Code: MSCCSMJE401**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

1. Understand the definition and significance of the Internet of Things
2. Discuss the architecture, operation, and business benefits of an IoT solution

3. Examine the potential business opportunities that IoT can uncover
4. Explore the relationship between IoT, cloud computing, and big data
5. Identify how IoT differs from traditional data collection systems

## Module 1:

Internet in general and Internet of Things: layers, protocols, packets, services, performance parameters of a packet network as well as applications such as web, Peer-to-peer, sensor networks, and multimedia. Transport services: TCP, UDP, socket programming. Network layer: forwarding & routing algorithms (Link, DV), IP-addresses, DNS, NAT, and routers. Local Area Networks, MAC level, link protocols such as: point-to-point protocols, Ethernet, WiFi 802.11, cellular Internet access, and Machine-to-machine. Mobile Networking: roaming and handoffs, mobile IP, and ad hoc and infrastructure less networks. Real-time networking: soft and real time, quality of service/information, resource reservation and scheduling, and performance measurements. IoT definitions: overview, applications, potential & challenges, and architecture.

**Text Books:**

1. Jan Holler, Vlasios Tsiatsis, Catherine Mulligan, Stefan Avesand, Stamatis Karnouskos, David Boyle, ―From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence‖, 1st Edition, Academic Press, 2014.
**2.** Vijay Madisetti and Arshdeep Bahga, ―Internet of Things (A Hands-on-Approach)‖, 1stEdition, VPT, 2014.

**Course Name: Green Computing**

**Course Code: MSCCSMJE402**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

Upon completion of the course, students should be able to:

1. give an account of the concept green IT and environmental perspectives on IT use
2. give an account of standards and certifications related to sustainable ITproducts and describe green IT in relation to technology,
3. relate green IT to sustainable development,
4. evaluate IT use in relation to environmental perspectives,
5. discuss how the choice of hardware and software can facilitate a moresustainable operation and to use methods and tools to measure energy consumption

## Module 1:

Logistics, Introduction to Green Computing & Background, Energy Management in Embedded Systems and Sensor Networks, Energy Management in Mobile Systems and Smartphones, Greening Desktop and Laptop PCs, Energy Efficient Networking and Communication, Greening Data Centers and Servers, IT Enabled Smart Buildings, Sensing within Buildings (Occupancy), Sensing within Buildings (Energy and Water), Managing the Data Deluge and ―App Platforms‖

for Smart Buildings, Energy Management in Smart Homes, Modeling, Prediction and Control for Smart Buildings, Security and Privacy.

**Text Books:**

**1.** The Green Computing Book: Tackling Energy Efficiency at Large Scale by Wu Chun Feng

**2.** Green Computing: Tools and Techniques for Saving Energy, Money, and Resources by Bud E. Smith

**Course Name: Cloud Computing**

**Course Code: MSCCSMJE403**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

Students will be able to

1. Understand the concepts, characteristics, delivery models and benefits of cloud computing
2. Understand the key security and compliance challenges of cloud computing
3. Understand the key technical and organisational challenges
4. Understand the different characteristics of public, private and hybrid cloud deployment models.

## Module 1:

Introduction: Cloud computing definition, reference model, Characteristics, Benefits, Challenges, Distributed Systems, Virtualization, Service-oriented computing, Utility-oriented computing, Overview on computing platforms & technologies – AWS, Google AppEngine, MS Azure, Hadoop, Saleforce.com, Manjrasoft Aneka Parallel & Distributed Computing: Parallel vs. Distributed computing, Elements of parallel computing, Parallel processing - hardware architecture & approaches, Concept & Component of Distributed Computing, RPC, Service-oriented computing Virtualization: Cloud reference model – IaaS, PaaS, SaaS, Types of clouds – Public, Private, Hybrid, Community, Cloud interoperability & standards, scalability & fault tolerance, Security, trust & privacy Concurrent Computing, High-throughput Computing and Data-Intensive Computing: Programming applications with Threads, Thread API, Parallel computation with Threads, Task computing, Frameworks for Task computing, Task-based application model, Data-intensive computing, characteristics, technology. Cloud Platforms and Applications: Overview on Amazon Web Services, Google AppEngine and Microsoft Azure, Cloud applications in scientific, business and consumer Domain

**Text Books:**

1. Buyya, Vecciola and Selvi, Mastering Cloud Computing: Foundations and Applications Programming, Tata McGraw Hill
2. Rittinghouse and Ransome, Cloud Computing: Implementation, Management, and Security, CRC Press
3. Aravind Doss, Cloud Computing, Tata McGraw Hill

4. Kris Jamsa, Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and More, Jones & Bartlett Learning

**Course Name: Computer Vision**

**Course Code: MSCCSMJE404**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

Students will be able to

1. Analyse and design a range of algorithms for image processing and computer vision
2. Develop and evaluate solutions to problems in computer vision
3. Demonstrate awareness of the current key research issues in computer vision

## Module 1:

Introduction : Image Processing, Computer Vision and Computer Graphics , What is Computer Vision - Low-level, Mid-level, High-level , Overview of Diverse Computer Vision Applications: Document Image Analysis, Biometrics, Object Recognition, Tracking, Medical Image Analysis, Content-Based Image Retrieval, Video Data Processing, Multimedia, Virtual Reality and Augmented Reality

Image Formation Models : Monocular imaging system , Radiosity: The 'Physics' of Image Formation, Radiance, Irradiance, BRDF, color etc, Orthographic & Perspective Projection,• Camera model and Camera calibration, Binocular imaging systems, Multiple views geometry, Structure determination, shape from shading , Photometric Stereo, Depth from Defocus , Construction of 3D model from images

Image Processing and Feature Extraction: Image pre-processing, Image representations (continuous and discrete), Edge detection

Motion Estimation: Regularization theory, Optical computation, Stereo Vision, Motion estimation, Structure from motion

Shape Representation and Segmentation: Contour based representation, Region based representation, Deformable curves and surfaces, Snakes and active contours, Level set representations, Fourier and wavelet descriptors, Medial representations, and Multiresolution analysis

Object recognition: Hough transforms and other simple object recognition methods, Shape correspondence and shape matching,Principal Component analysis, Shape priors for recognition

Image Understanding: Pattern recognition methods, HMM, GMM and EM

Applications: Photo album – Face detection – Face recognition – Eigen faces – Active appearance and 3D shape models of faces Application: Surveillance – foreground-background separation – particle filters – Chamfer matching, tracking, and occlusion – combining views from multiple cameras – human gait analysis Application: In-vehicle vision system: locating roadway – road markings – identifying road signs – locating pedestrians

**Text Books:**

1. Computer Vision - A modern approach, by D. Forsyth and J. Ponce, Prentice Hall Robot Vision, by B. K. P. Horn, McGraw-Hill.
2. Introductory Techniques for 3D Computer Vision, by E. Trucco and A. Verri, Publisher: Prentice Hall.
3. R. C. Gonzalez, R. E. Woods. Digital Image Processing. Addison Wesley Longman, Inc., 1992.
4. D. H. Ballard, C. M. Brown. Computer Vision. Prentice-Hall, Englewood Cliffs, 1982.
5. Richard Szeliski, Computer Vision: Algorithms and Applications (CVAA). Springer, 2010
6. Image Processing, Analysis, and Machine Vision. Sonka, Hlavac, and Boyle. Thomson.
7. E. R. Davies, Computer & Machine Vision, Fourth Edition, Academic Press, 2012
8. Simon J. D. Prince, Computer Vision: Models, Learning, and Inference, Cambridge University Press, 2012
9. Mark Nixon and Alberto S. Aquado, Feature Extraction & Image Processing for Computer Vision, Third Edition, Academic Press, 2012.

**Course Name: Deep Learning**

**Course Code: MSCCSMJE405**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***
Students will be able to

1. understand the fundamentals of deep learning.
2. know the main techniques in deep learning and the mainresearch in this field.
3. design and implement deep neural network systems.
4. identify new application requirements in the field ofcomputer vision.

## Module 1:

**Introduction**
Course logistics and overview.
**Linear Algebra Review**
Brief review of concepts from Linear Algebra.
**Optimization**
Types of errors, bias-variance trade-off, overfitting-underfitting, brief review of concepts from Vector Calculus and optimization, variants of gradient descent, momentum.
**Logistic Regression**
Basic concepts of rgression and classification problems, linear models addressing regression and classification, maximum likelihood, logistic regression classifiers.
**Neural Networks**
Basic concepts of artificial neurons, single and multi layer perceptrons, perceptron learning algorithm, its convergence proof, different activation functions, softmax cross entropy loss function.

**ConvNets**
Basic concepts of Convolutional Neural Networks starting from filetering. Convolution and pooling operation and arithmatics of these.
**ConvNet Architectures**
Discussions on famous convnet architectures - AlexNet, ZFNet, VGG, C3D, GoogLeNet, ResNet, MobileNet-v1.
**Regularization, Batchnorm**
Discussion on regularization, Dropout, Batchnorm etc.
**Detection, Segmentation (Part I)**
Discussion on detection, segmentation problem definition, challenges, Evaluation, Datasets and Localization by regression.
**Detection, Segmentation (Part II)**
Discussion on detection as classificaion, region proposals, RCNN architectures.
**Recurrent Neural Networks**
Discussion on Recurrent Neural Networks (RNNs), Long-Short Term Memory (LSTM) architectures and basics of word embedding.
**Vision and Language**
Discussion on different tasks involving Vision and Language e.g., Image and video captioning along with the use of attention.
**Explainability and Bias**
Discussion on explainability and bias in Deep Learning system. The need for explanation, introspection vs justification, activation maximization and activation map based explanation generation, Black-box explanation generation etc. Bias in AI and in image captioning task.

**Text Books:**

1. "Deep Learning", I Goodfellow, Y Bengio and A Courville, 1st Edition

## Course Name: Natural Language Processing

## Course Code: MSCCSMJE406

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***
Students will be able to

1. Understand approaches to syntax and semantics in NLP.
2. Understand approaches to discourse, generation, dialogue and summarization within NLP.
3. Understand current methods for statistical approaches to machine translation.
4. Understand machine learning techniques used in NLP, including hidden Markov models and probabilistic context-free grammars, clustering and unsupervised methods, log-linear and discriminative models, and the EM algorithm as applied within NLP

## Module I

Regular Expressions and Automata Recap
Introduction to NLP, Regular Expression, Finite State Automata
**Tokenization**
Word Tokenization, Normalization, Sentence Segmentation, Named Entity Recognition, Multi Word Extraction, Spell Checking – Bayesian Approach, Minimum Edit Distance
**Morphology**
Morphology – Inflectional and Derivational Morphology, Finite State Morphological Parsing, The Lexicon and
Morphotactics, Morphological Parsing with Finite State Transducers, Orthographic Rules and Finite State Transducers,
Porter Stemmer

## Module II

**Language Modeling**
Introduction to N-grams, Chain Rule, Smoothing – Add-One Smoothing, Witten-Bell Discounting; Backoff, Deleted
Interpolation, N-grams for Spelling and Word Prediction, Evaluation of language models.
**Hidden Markov Models and POS Tagging**
Markov Chain, Hidden Markov Models, Forward Algorithm, Viterbi Algorithm, Part of Speech Tagging – Rule based and
Machine Learning based approaches, Evaluation

## Module III

**Text Classification**
Text Classification, Naïve Bayes' Text Classification, Evaluation, Sentiment Analysis – Opinion Mining and Emotion
Analysis, Resources and Techniques
**Context Free Grammar**
Context Free Grammar and Constituency, Some common CFG phenomena for English, Top-Down and Bottom-up parsing,
Probabilistic Context Free Grammar, Dependency Parsing

## Module IV

**Computational Lexical Semantics**
Introduction to Lexical Semantics – Homonymy, Polysemy, Synonymy, Thesaurus – WordNet, Computational Lexical
Semantics – Thesaurus based and Distributional Word Similarity
**Information Retrieval**
Boolean Retrieval, Term-document incidence, The Inverted Index, Query Optimization, Phrase Queries, Ranked Retrieval
– Term Frequency – Inverse Document Frequency based ranking, Zone Indexing, Query term proximity, Cosine ranking,
Combining different features for ranking, Search Engine Evaluation, Relevance Feedback

**Books:**
1. Speech and Language Processing, Jurafsky and Martin, Pearson Education
2. Foundation of Statistical Natural Language Processing, Manning and Schutze, MIT Press

**Course Name: Fundamentals of Programming (C)**

**Course Code: MIE-1**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***

Students will be able to:

1. Implement the algorithms and draw flowcharts for solving Mathematical and Engineering problems.
2. Demonstrate an understanding of computer programming language concepts.
3. Develop C programs on linux platform.
4. Design and develop Computer programs, analyzes, and interprets the concept of pointers, declarations, initialization, operations on pointers and their usage.
5. Define data types and use them in simple data processing applications also he/she must be able to use the concept of array of structures. Student must be able to define union and enumeration user defined data types

# Module I

Unit 1:

Introduction to Programming

- Introduction to components of a computer system (disks, memory, processor, where a program is stored and executed, operating system, compilers etc.)
- Idea of Algorithm: steps to solve logical and numerical problems. Representation of Algorithm: Flowchart/Pseudocode with examples.
- From algorithms to programs; source code, variables (with data types) variables and memory locations, Syntax and Logical Errors in compilation, object and executable code

Unit 2: Arithmetic expressions and precedence

Unit 3: Conditional Branching and Loops

- Writing and evaluation of conditionals and consequent branching
- Iteration and loops

Unit 4: Arrays

- Arrays (1-D, 2-D), Character arrays and Strings

Unit 5: Basic Algorithms

- Searching, Basic Sorting Algorithms (Bubble, Insertion and Selection), Finding roots of equations, notion of order of complexity through example programs (no formal definition required)

Unit 6: Function

- Functions (including using built in libraries), Parameter passing in functions, call by value, Passing arrays to functions: idea of call by reference

Unit 7: Recursion

- Recursion, as a different way of solving problems. Example programs, such as Finding Factorial, Fibonacci series, Ackerman function etc.

Unit 8: Structure
• Structures, Defining structures and Array of Structures
Unit 9: Pointers
• Idea of pointers, Defining pointers, Use of Pointers in self-referential structures, notion of linked list (no implementation)
Unit 10: File handling

**Books:**
1. R. S. Salaria, Computer Concepts and Programming in C, Khanna Publishers
2. Byron Gottfried, Schaum's Outline of Programming with C, McGraw-Hill
3. E. Balaguruswamy, Programming in ANSI C, Tata McGraw-Hill
4. Brian W. Kernighan and Dennis M. Ritchie, The C Programming Language, Prentice Hall of India

**Course Name: Business Intelligence**

**Course Code: MIE-2**

| Course Type: **Theory** | Course Details: Core | | L-T-P: **4-0-0** |
|---|---|---|---|
| Credit: 4 | Full Marks: 50 | CA Marks | ESE Marks |
| | | Theoretical | Theoretical |
| | | **10** | **40** |

***Course Learning Outcomes:***
Students will be able to:

1. Describe the concepts and components of Business Intelligence (BI).
2. Critically evaluate use of BI for supporting decision making in an organisation.
3. Understand and use the technologies and tools that make up BI (e.g. Data warehousing, Data reporting and use of Online analytical processing (OLAP)).
4. Understand and design the technological architecture that underpins BI systems.
5. Plan the implementation of a BI system.

## Module I

UNIT I : BUSINESS INTELLIGENCE
Effective and timely decisions – Data, information and knowledge – Role of mathematical models – Business intelligence architectures: Cycle of a business intelligence analysis – Enabling factors in business intelligence projects – Development of a business intelligence system – Ethics and business intelligence.
UNIT II : KNOWLEDGE DELIVERY
The business intelligence user types, Standard reports, Interactive Analysis and Ad Hoc Querying, Parameterized Reports and Self-Service Reporting, dimensional analysis, Alerts/Notifications, Visualization: Charts, Graphs, Widgets, Scorecards and Dashboards, Geographic Visualization, Integrated Analytics, Considerations: Optimizing the Presentation for the Right Message.
UNIT III : EFFICIENCY
Efficiency measures – The CCR model: Definition of target objectives- Peer groups –

Identification of good operating practices; cross efficiency analysis – virtual inputs and outputs – Other models. Pattern matching – cluster analysis, outlier analysis

**TEXT BOOK:**


1. Efraim Turban, Ramesh Sharda, Dursun Delen, "Decision Support and Business Intelligence Systems", 9th Edition, Pearson 2013.
2. Larissa T. Moss, S. Atre, "Business Intelligence Roadmap: The Complete Project Lifecycle of Decision Making", Addison Wesley, 2003.
3. Carlo Vercellis, "Business Intelligence: Data Mining and Optimization for Decision Making", Wiley Publications, 2009.
4. David Loshin Morgan, Kaufman, "Business Intelligence: The Savvy Manager"s Guide", Second Edition, 2012.
5. Cindi Howson, "Successful Business Intelligence: Secrets to Making BI a Killer App", McGraw-Hill, 2007.
6. Ralph Kimball , Margy Ross , Warren Thornthwaite, Joy Mundy, Bob Becker, "The Data Warehouse Lifecycle Toolkit", Wiley Publication Inc.,2007.