MAHARAJA SUHEL DEV UNIVERSITY AZAMGARH

BACHELOR OF ENGINEERING (B.E.) DEGREE PROGRAMME

(CHOICE BASED CREDIT SYSTEM)

COURSE STRUCTURE AND EVALUATION SCHEME

B.E. COMPUTER SCIENCE ENGINEERING
(Data Science)

(4 YEARS UG PROGRAMME)

SYLLABI

For
B.E. 3rd Year
(Data Science)

Effective from the Session: 2026-27

B.E. Third Year Semester - V

SN	Subject Code	Subject	L	T	P	CT	TA	CT+TA	TE/PE	Total	Credit
1	BCS501	Database Management System	3	1	0	15	10	25	75	100	4
2	BCDS501	Introduction to Data Analytics and Visualization	3	1	0	15	10	25	75	100	4
3	BCS503	Design and Analysis of Algorithm	3	1	0	15	10	25	75	100	4
4	Deptt-Elective-I	Departmental Elective-I	3	0	0	15	10	25	75	100	3
5	Deptt-Elective-II	Departmental Elective-II	3	0	0	15	10	25	75	100	3
6	BCS551	Database Management System Lab	0	0	2	0	25	25	75	100	1
7	BCDS551	Data Analytics and Visualization Lab	0	0	2	0	25	25	75	100	1
8	BCS553	Design and Analysis of Algorithm Lab	0	0	2	0	25	25	75	100	1
9	BCS554	Mini Project / Internship Assessment	0	0	2	0	100	100	-	100	1
10	BNC501	Constitution of India	2	0	0	15	10	25	75	100	0
		Total	17	3	8						22

*The Mini Project or Internship (4 weeks) conducted during summer break after IV semester will be assessed during V semester.

Maharaja Suhel Dev University, Azamgarh

B.E. Third Year Semester - VI

SN.	Subject Code	Subject	L	T	P	CT	TA	CT+TA	TE/PE	Total	Credit
1	BCS601	Software Engineering	3	1	0	15	10	25	75	100	4
2	BCDS601	Big Data and Analytics	3	1	0	15	10	25	75	100	4
3	BCS603	Computer Networks	3	1	0	15	10	25	75	100	4
4	Deptt-Elective-III	Departmental Elective-III	3	0	0	15	10	25	75	100	3
5	Open Elective-I	Open Elective-I	3	0	0	15	10	25	75	100	4
6	BCS651	Software Engineering Lab	0	0	2	0	25	25	75	100	1
7	BCDS651	Big Data and Analytics Lab	0	0	2	0	25	25	75	100	1
8	BCS653	Computer Networks Lab	0	0	2	0	25	25	75	100	1
9	BNC602	Essence of Indian Traditional Knowledge	2	0	0	15	10	25	75	100	0
		Total	**17**	**3**	**6**					**900**	**22**

* The Mini Project or Internship (4 weeks) will be done during summer break after VI Semester and will be assessed during VII semester.

Departmental Elective-I

1. BCAI051 - Mathematical Foundation AI, ML and Data Science
2. BCS058 - Data Warehouse & Data Mining
3. BCDS051 - Business Intelligence and Analytics
4. BCS054 - Object Oriented System Design with C++

Departmental Elective-II

1. BCAM051 - Cloud Computing
2. BCAI052 - Natural Language Processing
3. BCS056 - Application of Soft Computing
4. BCS057- Image Processing

Departmental Elective-III

1. BCAI061 - Cyber Forensic analytics
2. BCDS061 - Image Analytics
3. BCAM061 - Social Media Analytics and Data Analysis
4. BCDS062 – Machine Learning Techniques

DATABASE MANAGEMENT SYSTEM (BCS501)

Unit I: Introduction to Database Systems

Overview of Database Management Systems, Difference between Database System and File System, Database System Concepts and Architecture, Data Models, Schema and Instances, Data Independence, Database Languages and Interfaces, Data Definition Language (DDL), Data Manipulation Language (DML), Overall Database Structure.

Unit II: Entity Relationship Model and Relational Model

ER Model Concepts, ER Diagram Notations, Mapping Constraints, Keys (Super Key, Candidate Key, Primary Key), Generalization and Aggregation, Conversion of ER Diagrams into Tables, Extended ER Model, Higher Degree Relationships.

Relational Data Model Concepts, Integrity Constraints, Entity Integrity, Referential Integrity, Domain Constraints, Relational Algebra, Relational Calculus, Tuple Relational Calculus and Domain Relational Calculus.

Unit III: SQL and PL/SQL

Introduction to SQL, Characteristics and Advantages of SQL, SQL Data Types and Literals, Types of SQL Commands, SQL Operators, Tables, Views and Indexes, Queries and Subqueries, Aggregate Functions, Insert, Update and Delete Operations, Joins, Union, Intersection and Minus Operations, Cursors, Triggers and Procedures in SQL/PL SQL.

Unit IV: Database Design and Transaction Processing

Functional Dependency and Normalization, First Normal Form (1NF), Second Normal Form (2NF), Third Normal Form (3NF), BCNF, Inclusion Dependency, Lossless Join Decomposition, Normalization using Functional Dependency (FD), Multivalued Dependency (MVD) and Join Dependency (JD), Alternative Approaches to Database Design.

Transaction Processing Concepts, Transaction Systems, Serializability of Schedules, Conflict and View Serializable Schedules, Recoverability, Recovery from Transaction Failures, Log-Based Recovery, Checkpoints and Deadlock Handling.

Unit V: Distributed Database and Concurrency Control

Distributed Database Concepts, Distributed Data Storage, Concurrency Control in Distributed Databases, Directory Systems.

Concurrency Control Techniques, Locking Techniques, Time Stamping Protocols, Validation-Based Protocols, Multiple Granularity, Multi-Version Schemes, Recovery with Concurrent Transactions, Case Study of Oracle Database.

Text Books

1. Database System Concepts — Abraham Silberschatz, Henry F. Korth and S. Sudarshan, McGraw Hill.
2. An Introduction to Database Systems — C. J. Date, Addison Wesley.
3. Fundamentals of Database Systems — Ramez Elmasri and Shamkant B. Navathe, Addison Wesley.
4. Databases — Patrick O'Neil, Elsevier Publications.
5. Database Management Systems — Raghu Ramakrishnan, McGraw Hill.
6. Database Management Systems — Alexis Leon and Mathews Leon, Vikas Publishing House.
7. An Introduction to Database Systems — Bipin C. Desai, Gagotia Publications.
8. Database Management System — Nilanjan Majumdar and Pritimoy Bhattacharya, Tata McGraw Hill (TMH).

Introduction to Data Analytics and Visualization (BAI072)

Unit I: Introduction to Data Analytics

Introduction to Data Analytics, sources and nature of data, classification of data including structured, semi-structured, and unstructured data, characteristics of data, introduction to Big Data platforms, need for data analytics, evolution of analytic scalability, analytic processes and tools, analysis versus reporting, modern data analytics tools, and applications of data analytics.

Data Analytics Lifecycle: Need for analytics, key roles in successful analytics projects, and phases of the data analytics lifecycle including discovery, data preparation, model planning, model building, communicating results, and operationalization.

Unit II: Data Analysis Techniques

Regression modeling, multivariate analysis, Bayesian modeling, inference and Bayesian networks, support vector machines and kernel methods.

Analysis of time series including linear systems analysis and nonlinear dynamics, rule induction, neural networks, learning and generalization, competitive learning, principal component analysis (PCA) and neural networks.

Fuzzy logic concepts, extracting fuzzy models from data, fuzzy decision trees, and stochastic search methods.

Unit III: Mining Data Streams

Introduction to stream concepts, stream data models and architecture, stream computing, sampling data in streams, filtering streams, counting distinct elements in streams, estimating moments, counting ones in a window, and decaying windows.

Real-Time Analytics Platforms (RTAP) and their applications. Case studies including real-time sentiment analysis and stock market prediction.

Unit IV: Frequent Itemsets and Clustering

Mining frequent itemsets, market-based modeling, Apriori algorithm, handling large datasets in main memory, limited-pass algorithms, and counting frequent itemsets in data streams.

Clustering techniques including hierarchical clustering, K-means clustering, clustering of high-dimensional data, CLIQUE and ProCLUS algorithms, frequent pattern-based clustering methods, clustering in non-Euclidean spaces, and clustering for streams and parallel environments.

Unit V: Data Visualization

Introduction to visualization, stages of visualization, computational support, visualization issues, different types of tasks, data representation techniques, limitations of display space, rendering time, and navigation links.

Human vision and perception, spatial and temporal limitations, visualization design principles, exploration of complex information spaces, figure captions in visual interfaces, visual objects and data objects, space perception and data representation in space.

Use of images, narratives, and gestures for effective explanation and communication of information.

Text Books

1. S. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach", Prentice Hall, Third Edition, 2009.
2. I. Bratko, "Prolog: Programming for Artificial Intelligence", Fourth Edition, Addison-Wesley Educational Publishers Inc., 2011.
3. M. Tim Jones, "Artificial Intelligence: A Systems Approach", Jones and Bartlett Publishers Inc., First Edition, 2008.
4. Nils J. Nilsson, "The Quest for Artificial Intelligence", Cambridge University Press, 2009.
5. William F. Clocksin and Christopher S. Mellish, "Programming in Prolog: Using the ISO Standard", Fifth Edition, Springer, 2003.
6. Gerhard Weiss, "Multi Agent Systems", Second Edition, MIT Press, 2013.
7. David L. Poole and Alan K. Mackworth, "Artificial Intelligence: Foundations of Computational Agents", Cambridge University Press, 2010.

DESIGN AND ANALYSIS OF ALGORITHM (BCS503)

Unit I: Introduction and Sorting Algorithms

Introduction to Algorithms, Analysis of Algorithms, Time and Space Complexity, Growth of Functions, Performance Measurement of Algorithms.

Sorting and Order Statistics: Shell Sort, Quick Sort, Merge Sort, Heap Sort, Comparison of Sorting Algorithms, Linear Time Sorting Techniques.

Unit II: Advanced Data Structures and Divide & Conquer

Advanced Data Structures: Red-Black Trees, B-Trees, Binomial Heaps, Fibonacci Heaps, Tries and Skip Lists.

Divide and Conquer Technique: Concept and Applications in Sorting, Matrix Multiplication, Convex Hull and Searching Problems.

Unit III: Greedy and Dynamic Programming Techniques

Greedy Method: Knapsack Problem, Optimal Reliability Allocation, Minimum Spanning Trees using Prim's and Kruskal's Algorithms, Single Source Shortest Path using Dijkstra's and Bellman-Ford Algorithms.

Dynamic Programming: Knapsack Problem, All-Pairs Shortest Paths using Warshall's and Floyd's Algorithms, Resource Allocation Problems.

Unit IV: Backtracking and Branch & Bound

Backtracking Technique: Travelling Salesman Problem, Graph Coloring, N-Queen Problem, Hamiltonian Cycles and Sum of Subsets Problem.

Branch and Bound Method: Concept and Applications in Optimization Problems.

Unit V: Advanced Algorithm Topics

Algebraic Computation, Fast Fourier Transform (FFT), String Matching Algorithms, Theory of NP-Completeness, Approximation Algorithms and Randomized Algorithms.

Text Books

1. Introduction to Algorithms — Thomas H. Cormen, Charles E. Leiserson and Ronald L. Rivest, Prentice Hall of India.
2. Fundamentals of Computer Algorithms — Ellis Horowitz and Sartaj Sahni.
3. The Design and Analysis of Computer Algorithms — Alfred V. Aho, John Hopcroft and Jeffrey Ullman, Pearson Education, 2008.
4. Design and Analysis of Algorithms — McGraw Hill.
5. Foundations of Algorithms — Richard E. Neapolitan, Jones & Bartlett Learning.
6. Algorithm Design — Jon Kleinberg and Éva Tardos, Pearson, 2005.
7. Algorithm Design: Foundations, Analysis, and Internet Examples — Michael T. Goodrich and Roberto Tamassia, Wiley, 2006.
8. Data Structures and Their Algorithms — Harry R. Lewis and Larry Denenberg, Harper Collins, 1997.
9. Algorithms — Robert Sedgewick and Kevin Wayne, Addison Wesley, 2011.
10. Algorithm Design and Analysis — Harsh Bhasin, Oxford University Press.

Departmental Elective-I

MATHEMATICAL FOUNDATION FOR AI, ML AND DATA SCIENCE(BCAI051)

Unit I – Descriptive Statistics and Probability

Descriptive Statistics: Diagrammatic representation of data, measures of central tendency, measures of dispersion, measures of skewness and kurtosis, correlation, inference procedure for correlation coefficient, bivariate correlation, multiple correlations, linear regression and its inference procedure, multiple regression.

Probability: Measures of probability, conditional probability, independent events, Bayes' theorem, random variables, discrete and continuous probability distributions, expectation and variance, Markov inequality, Chebyshev's inequality and Central Limit Theorem.

Unit II – Inferential Statistics and Regression Analysis

Inferential Statistics: Sampling and confidence interval, inference and significance, estimation and hypothesis testing, goodness of fit, test of independence, permutations and randomization test, t-test/z-test including one sample, independent and paired tests, ANOVA and chi-square test.

Linear Methods for Regression Analysis: Multiple regression analysis, orthogonalization using Householder transformations (QR), singular value decomposition (SVD) and linear dimension reduction using principal component analysis (PCA).

Unit III – Random Numbers and Monte Carlo Methods

Pseudo-Random Numbers: Random number generation, inverse-transform method, acceptance-rejection method, transformations and multivariate probability calculations.

Monte Carlo Integration: Simulation and Monte Carlo integration, variance reduction, Monte Carlo hypothesis testing, antithetic variables, control variates, importance sampling and stratified sampling.

Markov Chain Monte Carlo (MCMC): Markov chains, Metropolis-Hastings algorithm, Gibbs sampling and convergence.

Unit IV – Vector Spaces and Inner Product Spaces

Vector Spaces: Vector space, subspace, linear combination, linear independence, basis, dimension, finding a basis of a vector space, coordinates and change of basis.

Inner Product Spaces: Inner product, length, orthogonal vectors, triangle inequality, Cauchy-Schwarz inequality, orthonormal basis and Gram-Schmidt process.

Unit V – Linear Transformations and Eigenvalues

Linear Transformations: Linear transformations and matrices for linear transformation, kernel and range of linear transformations and change of basis.

Eigenvalues and Eigenvectors: Definition of eigenvalue and eigenvector, diagonalization, symmetric matrices and orthogonal diagonalization.

References

1. S.C. Gupta and V.K. Kapoor, "Fundamentals of Mathematical Statistics", Sultan Chand & Sons.
2. Sheldon M. Ross, "Introduction to Probability and Statistics for Engineers and Scientists", Academic Press.
3. Dudewicz, E.J. and Mishra, S.N., "Modern Mathematical Statistics", Wiley.
4. Purohit S.G., Gore S.D. and Deshmukh S.K., "Statistics Using R", Narosa.
5. Rizzo, M.L., "Statistical Computing with R", Chapman & Hall/CRC Press.
6. Norman Matloff, "The Art of R Programming", No Starch Press.
7. Dalgaard, Peter, "Introductory Statistics with R", Springer Science & Business Media.
8. M.D. Ugarte, A.F. Militino and A.T. Arnholt, "Probability and Statistics with R", CRC Press.
9. Kundu, D. and Basu, A., "Statistical Computing – Existing Methods and Recent Developments", Narosa.
10. Gentle, James E., Härdle, Wolfgang Karl and Mori, Yuichi, "Handbook of Computational Statistics", Springer.
11. Givens and Hoeting, "Computational Statistics", Wiley Series in Probability and Statistics.
12. Ron Larson, "Elementary Linear Algebra", 8th Edition, Cengage Learning, 2017.

Departmental Elective-I

DATA WAREHOUSING AND DATA MINING (BCS058)

Unit I – Data Warehousing

Overview, definition, components of data warehousing, building a data warehouse, warehouse database, mapping the data warehouse to a multiprocessor architecture, difference between database systems and data warehouses, multidimensional data model, data cubes, stars, snowflakes and fact constellation concepts.

Unit II – Data Warehouse Process and Technology

Warehousing strategy, warehouse management and support processes, warehouse planning and implementation, hardware and operating systems for data warehousing, client/server computing model and data warehousing, parallel processors and cluster systems, distributed DBMS implementation, warehousing software and warehouse schema design.

Unit III – Data Mining

Overview, motivation, definition and functionalities, data processing, forms of data preprocessing, data cleaning, missing values, noisy data, binning, clustering, regression, computer and human inspection, inconsistent data, data integration and transformation.

Data Reduction: Data cube aggregation, dimensionality reduction, data compression, numerosity reduction, discretization, concept hierarchy generation and decision tree.

Unit IV – Classification and Clustering

Classification: Definition, data generalization, analytical characterization, analysis of attribute relevance, mining class comparisons, statistical measures in large databases, statistical-based algorithms, distance-based algorithms and decision tree-based algorithms.

Clustering: Introduction, similarity and distance measures, hierarchical and partitional algorithms, hierarchical clustering using CURE and Chameleon, density-based methods such as DBSCAN and OPTICS, grid-based methods including STING and CLIQUE, model-based statistical approaches.

Association Rules: Introduction, large item sets, basic algorithms, parallel and distributed algorithms and neural network approaches.

Unit V – Data Visualization and Applications

Data visualization and overall perspective, aggregation, historical information, query facility, OLAP functions and tools, OLAP servers, ROLAP, MOLAP, HOLAP, data mining interface, security, backup and recovery, tuning and testing of data warehouses.

Warehousing applications and recent trends including web mining, spatial mining and temporal mining.

Text Books

1. Alex Berson and Stephen J. Smith, "Data Warehousing, Data-Mining & OLAP", TMH.
2. Mark Humphries, Michael W. Hawkins and Michelle C. Dy, "Data Warehousing: Architecture and Implementation", Pearson.
3. Margaret H. Dunham and S. Sridhar, "Data Mining: Introductory and Advanced Topics", Pearson Education.
4. Arun K. Pujari, "Data Mining Techniques", Universities Press.
5. Pieter Adriaans and Dolf Zantinge, "Data-Mining", Pearson Education.

Maharaja Suhel Dev University, Azamgarh

Departmental Elective-I

DATA ANALYTICS(BCS052)

Unit I – Introduction to Data Analytics

Sources and nature of data, classification of data including structured, semi-structured and unstructured data, characteristics of data, introduction to Big Data platform, need for data analytics, evolution of analytic scalability, analytic process and tools, analysis vs reporting, modern data analytic tools and applications of data analytics.

Data Analytics Lifecycle: Need, key roles for successful analytic projects, phases of data analytics lifecycle including discovery, data preparation, model planning, model building, communicating results and operationalization.

Unit II – Data Analysis

Regression modeling, multivariate analysis, Bayesian modeling, inference and Bayesian networks, support vector and kernel methods, analysis of time series including linear systems analysis and nonlinear dynamics, rule induction, neural networks, learning and generalization, competitive learning, principal component analysis and neural networks, fuzzy logic, extraction of fuzzy models from data, fuzzy decision trees and stochastic search methods.

Unit III – Mining Data Streams

Introduction to stream concepts, stream data model and architecture, stream computing, sampling data in a stream, filtering streams, counting distinct elements in a stream, estimating moments, counting ones in a window, decaying window, Real-Time Analytics Platform (RTAP) applications, case studies including real-time sentiment analysis and stock market prediction.

Unit IV – Frequent Item Sets and Clustering

Mining frequent item sets, market-based modeling, Apriori algorithm, handling large datasets in main memory, limited pass algorithm, counting frequent item sets in a stream.

Clustering Techniques: Hierarchical clustering, K-means clustering, clustering of high-dimensional data, CLIQUE and ProCLUS, frequent pattern-based clustering methods, clustering in non-Euclidean space, clustering for streams and parallelism.

Unit V – Frameworks and Visualization

Frameworks: MapReduce, Hadoop, Pig, Hive, HBase, MapR, sharding, NoSQL databases, S3 and Hadoop Distributed File Systems.

Visualization: Visual data analysis techniques, interaction techniques, systems and applications.

Introduction to R: R graphical user interfaces, data import and export, attributes and data types, descriptive statistics, exploratory data analysis, visualization before analysis and analytics for unstructured data.

Text books and References:

1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
2. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press.
3.Bill Franks, Taming the Big Data Tidal wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley & Sons.
3. John Garrett, Data Analytics for IT Networks : Developing Innovative Use Cases, Pearson Education
4. Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley
5.David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big Data Analytics", EMC Education Series, John Wiley
6. Frank J Ohlhorst, "Big Data Analytics: Turning Big Data into Big Money", Wiley and SAS Business Series
7. Colleen Mccue, "Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis", Elsevier
8. Michael Berthold, David J. Hand," Intelligent Data Analysis", Springer

Departmental Elective-I

OBJECT ORIENTED SYSTEM DESIGN WITH C++ (BSC054)

Unit I – Introduction to Object Orientation

The meaning of object orientation, object identity, encapsulation, information hiding, polymorphism, generality, importance of modelling, principles of modelling, object-oriented modelling, introduction to UML, conceptual model of UML and architecture.

Unit II – Structural and Behavioural Modeling

Basic Structural Modeling: Classes, relationships, common mechanisms and diagrams. Class and object diagrams, terms, concepts and modelling techniques for class and object diagrams.
Collaboration Diagrams: Terms, concepts, depicting messages, polymorphism in collaboration diagrams, iterated messages and use of self in messages.
Sequence Diagrams: Terms, concepts, depicting asynchronous messages with or without priority, call-back mechanisms and broadcast messages.
Basic Behavioural Modeling: Use cases, use case diagrams, activity diagrams, state machines, processes and threads, events and signals, time diagrams, interaction diagrams and package diagrams.
Architectural Modeling: Components, deployment, component diagrams and deployment diagrams.

Unit III – Object Oriented Analysis and Design

Object oriented analysis, object-oriented design, combining three models, designing algorithms, design optimization, implementation of control, adjustment of inheritance, object representation, physical packaging and documenting design considerations.
Structured Analysis and Structured Design (SA/SD), Jackson Structured Development (JSD), mapping object-oriented concepts using non-object-oriented languages, translating classes into data structures, passing arguments to methods, implementing inheritance and association encapsulation.
Object oriented programming style including reusability, extensibility, robustness and programming in the large, procedural versus object-oriented programming, object-oriented language features, abstraction and encapsulation.

Unit IV – C++ Basics and Functions

C++ Basics: Overview, program structure, namespace, identifiers, variables, constants, enum, operators, typecasting and control structures.
C++ Functions: Simple functions, call and return by reference, inline functions, macro versus inline functions, function overloading, default arguments, friend functions and virtual functions.

Unit V – Objects, Classes and Polymorphism

Objects and Classes: Basics of objects and classes in C++, private and public members, static data and function members, constructors and their types, destructors, operator overloading and type conversion.
Inheritance: Concept of inheritance, types of inheritance including single, multiple, multilevel, hierarchical and hybrid inheritance, protected members, overriding and virtual base class.
Polymorphism: Pointers in C++, pointers and objects, this pointer, virtual and pure virtual functions and implementing polymorphism.

Text Books

1. James Rumbaugh et al., "Object Oriented Modeling and Design", 2nd Edition, Pearson Education.
2. Grady Booch, James Rumbaugh and Ivar Jacobson, "The Unified Modeling Language User Guide", Pearson Education.
3. E. Balagurusamy, "Object Oriented Programming with C++", McGraw-Hill Education.
4. Steven Holzner, "C++ Programming Black Book", Dreamtech.
5. Robert Lafore, "Object Oriented Programming in Turbo C++", Galgotia.
6. Ashok Kamthane, "Object Oriented Programming with ANSI and Turbo C++", Pearson.
7. Herbert Schildt, "The Complete Reference C++", McGraw-Hill Education.

Departmental Elective-II

CLOUD COMPUTING (BCAM051)

Unit I – Introduction to Cloud Computing

Definition of Cloud, Evolution of Cloud Computing, Principles of Parallel and Distributed Computing, Cloud Characteristics, Elasticity in Cloud, On-Demand Provisioning.

Unit II – Cloud Enabling Technologies & Service Oriented Architecture

REST and Systems of Systems, Web Services, Publish–Subscribe Model, Basics of Virtualization, Types of Virtualization, Levels of Virtualization, Virtualization Structures, Tools and Mechanisms, CPU Virtualization, Memory Virtualization, I/O Virtualization, Virtualization Support, Disaster Recovery.

Unit III – Cloud Architecture, Services and Storage

Layered Cloud Architecture, NIST Cloud Computing Reference Architecture, Public Cloud, Private Cloud, Hybrid Cloud, IaaS, PaaS, SaaS, Cloud Design Challenges, Cloud Storage, Storage-as-a-Service, Advantages of Cloud Storage, Cloud Storage Providers, Amazon S3.

Unit IV – Resource Management and Security in Cloud

Inter-Cloud Resource Management, Resource Provisioning, Resource Provisioning Methods, Global Exchange of Cloud Resources, Security Overview, Cloud Security Challenges, SaaS Security, Security Governance, Virtual Machine Security, Identity and Access Management (IAM), Security Standards.

Unit V – Cloud Technologies and Advancements

Hadoop, MapReduce, VirtualBox, Google App Engine, Programming Environment for Google App Engine, OpenStack, Federation in Cloud Computing, Levels of Federation, Federated Services and Applications, Future of Federation.

Text Books

1. Kai Hwang, Geoffrey C. Fox, Jack G. Dongarra, *Distributed and Cloud Computing: From Parallel Processing to the Internet of Things*, Morgan Kaufmann Publishers, 2012.
2. John W. Rittinghouse and James F. Ransome, *Cloud Computing: Implementation, Management and Security*, CRC Press, 2017.
3. Rajkumar Buyya, Christian Vecchiola, and S. Thamarai Selvi, *Mastering Cloud Computing*, Tata McGraw Hill, 2013.
4. Toby Velte, Anthony Velte, and Robert Elsenpeter, *Cloud Computing: A Practical Approach*, Tata McGraw Hill, 2009.
5. George Reese, *Cloud Application Architectures: Building Applications and Infrastructure in the Cloud*, O'Reilly Media, 2009.

Departmental Elective-II

NATURAL LANGUAGE PROCESSING (BCAI052)

Unit I – Introduction to NLP

Origins and challenges of Natural Language Processing, language modelling including grammar-based and statistical language models, regular expressions, finite-state automata, English morphology, transducers for lexicon and rules, tokenization, detecting and correcting spelling errors, minimum edit distance.

Word Level Analysis: Unsmoothed N-grams, evaluating N-grams, smoothing, interpolation and backoff, word classes, word tokenization, TF-IDF vectors, semantic analysis, outline of English syntax, introduction to semantics and knowledge representation and Zipf's Law.

Unit II – Syntactic Analysis

Context free grammars, grammar rules for English, grammars and parsing, sentence structure, top-down and bottom-up parsers, transition network grammars and top-down chart parsing.

Feature Systems and Augmented Grammars: Basic feature systems for English, morphological analysis and lexicon, parsing with features and augmented transition networks.

Ambiguity Resolution: Statistical methods, probabilistic language processing, estimating probabilities, part-of-speech tagging, obtaining lexical probabilities, probabilistic context-free grammars, best-first parsing, feature structures and unification of feature structures.

Unit III – Semantics and Pragmatics

Requirements for representation, first-order logic, description logics, syntax-driven semantic analysis, semantic attachments, word senses, relations between senses, thematic roles, selection restrictions and word sense disambiguation.

Word Sense Disambiguation using supervised methods, dictionary and thesaurus methods, bootstrapping methods and word similarity using thesaurus and distributional methods.

Unit IV – Basic Concepts of Speech Processing

Speech fundamentals, articulatory phonetics, production and classification of speech sounds, acoustic phonetics and acoustics of speech production.

Review of digital signal processing concepts, Short-Time Fourier Transform (STFT), Filter Bank methods and Linear Predictive Coding (LPC) methods.

Unit V – Speech Analysis and NLP Applications

Speech analysis features, feature extraction and pattern comparison techniques, speech distortion measures including mathematical and perceptual measures.

Real-world NLP challenges including information extraction, question answering systems, dialog engines, optimization, parallelization and batch processing.

Text books:

- Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication, 2014.
- Steven Bird, Ewan Klein and Edward Loper, —Natural Language Processing with Python, First Edition, OReilly Media, 2009.
- Lawrence Rabiner And Biing-Hwang Juang, "Fundamentals Of Speech Recognition", Pearson Education, 2003. Daniel Jurafsky And James H Martin, "Speech And Language Processing – An Introduction To Natural Language Processing, Computational Linguistics, And Speech Recognition", Pearson Education, 2002.
- Frederick Jelinek, "Statistical Methods Of Speech Recognition", MIT Press, 1997.
- Breck Baldwin, —Language Processing with Java and LingPipe Cookbook, Atlantic Publisher, 2015. Richard M Reese, — Natural Language Processing with Java, OReilly Media, 2015.
- Nitin Indurkhya and Fred J. Damerau, —Handbook of Natural Language Processing, Second Edition, Chapman and Hall/CRC Press, 2010.

Departmental Elective-II

APPLICATION OF SOFT COMPUTING (BCS056)

Unit I: Neural Networks – I (Introduction and Architecture)

Introduction to Neural Networks, Biological Neuron, Nerve Structure and Synapse, Artificial Neuron Model, Activation Functions.

Neural Network Architectures: Single Layer and Multilayer Feed Forward Networks, Recurrent Networks. Learning Techniques: Perceptron Learning Rule, Convergence Rule, Auto-Associative Memory and Hetero-Associative Memory.

Unit II: Neural Networks – II (Backpropagation Networks)

Perceptron Model and its Solution, Single Layer Artificial Neural Networks, Multilayer Perceptron Model. Backpropagation Networks: Learning Methods, Learning Rule Coefficient, Backpropagation Algorithm, Factors Affecting Backpropagation Training and Applications of Neural Networks.

Unit III: Fuzzy Logic – I (Introduction)

Basic Concepts of Fuzzy Logic, Fuzzy Sets and Crisp Sets, Fuzzy Set Theory and Operations, Properties of Fuzzy Sets.

Fuzzy Relations and Crisp Relations, Conversion of Fuzzy Data to Crisp Data.

Unit IV: Fuzzy Logic – II (Membership Functions and Rules)

Membership Functions, Inference Mechanism in Fuzzy Logic, Fuzzy IF-THEN Rules, Fuzzy Implications and Fuzzy Algorithms.

Fuzzification and Defuzzification Techniques, Fuzzy Controllers and Industrial Applications of Fuzzy Logic.

Unit V: Genetic Algorithms (GA)

Introduction to Genetic Algorithms, Basic Concepts and Working Principle of GA, Flowchart and Procedure of Genetic Algorithms.

Genetic Representation and Encoding, Initialization and Selection Process, Genetic Operators, Mutation, Generational Cycle and Applications of Genetic Algorithms.

Text Books

1. Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis and Applications — S. Rajasekaran and G. A. Vijayalakshmi Pai, Prentice Hall of India.
2. Artificial Intelligence and Intelligent Systems — N. P. Padhy, Oxford University Press.

Reference Books

3. Neural Networks — Simon Haykin, Pearson Education.
4. Fuzzy Logic with Engineering Applications — Timothy J. Ross, Wiley India.
5. Neural Networks — Satish Kumar, McGraw Hill.

Maharaja Suhel Dev University, Azamgarh

Departmental Elective-II

IMAGE PROCESSING (BCS057)

Unit I: Digital Image Fundamentals

Introduction to Digital Image Processing, Steps in Digital Image Processing, Components of Image Processing Systems, Elements of Visual Perception, Image Sensing and Acquisition, Image Sampling and Quantization, Relationship between Pixels.

Color Image Fundamentals: RGB and HSI Color Models.

Two-Dimensional Mathematical Transforms: Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT).

Unit II: Image Enhancement

Spatial Domain Enhancement

Gray Level Transformations, Histogram Processing, Basics of Spatial Filtering, Smoothing and Sharpening Filters.

Frequency Domain Enhancement

Introduction to Fourier Transform, Frequency Domain Smoothing and Sharpening Filters, Ideal Filters, Butterworth Filters and Gaussian Filters, Homomorphic Filtering and Color Image Enhancement Techniques.

Unit III: Image Restoration

Image Degradation and Restoration Model, Noise Models, Mean Filters, Order Statistics Filters, Adaptive Filters.

Frequency Domain Restoration: Band Reject Filters, Band Pass Filters, Notch Filters, Optimum Notch Filtering.

Image Restoration Techniques: Inverse Filtering and Wiener Filtering.

Unit IV: Image Segmentation

Edge Detection Techniques, Edge Linking using Hough Transform, Thresholding Methods.

Region-Based Segmentation: Region Growing, Region Splitting and Merging.

Morphological Processing: Erosion and Dilation.

Watershed Segmentation: Basic Concepts, Dam Construction and Watershed Segmentation Algorithm.

Unit V: Image Compression and Recognition

Need for Image Compression, Huffman Coding, Run Length Encoding, Shift Codes, Arithmetic Coding, JPEG and MPEG Standards.

Image Representation and Recognition: Boundary Representation, Boundary Description, Fourier Descriptors, Regional Descriptors, Topological Features, Texture Analysis, Pattern Recognition and Recognition based on Matching Techniques.

Text Books

1. Digital Image Processing — Rafael C. Gonzalez and Richard E. Woods, Pearson, 2010.
2. Fundamentals of Digital Image Processing — Anil K. Jain, Pearson, 2002.
3. Digital Image Processing — Kenneth R. Castleman, Pearson, 2006.
4. Digital Image Processing Using MATLAB — Steven Eddins, Pearson Education, 2011.
5. Multidimensional Digital Signal Processing — Dan E. Dudgeon and Russell M. Mersereau, Prentice Hall, 1990.
6. Digital Image Processing — William K. Pratt, John Wiley & Sons, 2002.

DATABASE MANAGEMENT SYSTEMS LAB (BCS551)

B.E. Computer Science Engineering (Data Science) – Third Year (2026-27)

Detailed Syllabus

Experiment 1: Installation of Database Software

Installation and Configuration of Oracle/MySQL Database Server.

Experiment 2: Entity Relationship Diagram

Creating Entity-Relationship (ER) Diagrams using CASE Tools.

Experiment 3: SQL Programming using Oracle/MySQL

- Writing Basic SQL SELECT Statements
- Restricting and Sorting Data
- Retrieving Data from Multiple Tables
- Using Group Functions and Aggregate Functions
- Data Manipulation Operations (INSERT, UPDATE, DELETE)
- Creating and Managing Tables

Experiment 4: Database Normalization

Normalization Techniques: 1NF, 2NF, 3NF and BCNF.

Experiment 5: Cursors

Creating and Using Cursors in PL/SQL.

Experiment 6: Procedures and Functions

Creating Stored Procedures and Functions.

Experiment 7: Packages and Triggers

Creating and Managing Packages and Database Triggers.

Experiment 8: Payroll Processing System

Design and Implementation of Payroll Processing System.

Experiment 9: Library Information System

Design and Implementation of Library Information System.

Experiment 10: Student Information System

Design and Implementation of Student Information System.

Experiment 11: Backup and Recovery

Automatic Backup and Recovery of Database Files.

Experiment 12: Mini Project

Design and Development of Database Applications for any one of the following:

- Inventory Control System
- Material Requirement Processing System
- Hospital Management System
- Railway Reservation System
- Personal Information System
- Web-Based User Identification System
- Timetable Management System
- Hotel Management System

Note: The instructor may add, delete or modify experiments as required. Open-source database tools such as MySQL, SQL Server, Oracle, MongoDB, Cubrid and MariaDB are recommended for conducting the lab.

BCDS551: DATA ANALYTICS AND VISUALIZATION LAB

1. **Numerical Operations in R**
   - Write programs to accept user input and perform numerical operations such as MAX, MIN, AVG, SUM, SQRT, and ROUND functions using R.
2. **Data Import and Export**
   - Perform data import and export operations using different file formats such as CSV, XLS, and TXT with data frames in R.
3. **Matrix Operations Using R**
   - Create matrices using vectors and perform matrix addition, subtraction, multiplication, inverse, transpose, and division operations.
4. **Statistical Analysis**
   - Implement statistical measures such as Mean, Median, Mode, and Standard Deviation using R.
5. **Data Preprocessing**
   - Perform preprocessing operations including:
     - Handling Missing Values
     - Min-Max Normalization
6. **Dimensionality Reduction**
   - Apply Principal Component Analysis (PCA) on a housing dataset for dimensionality reduction and feature extraction.
7. **Simple Linear Regression**
   - Implement Simple Linear Regression in R and analyze prediction results.
8. **K-Means Clustering**
   - Perform K-Means clustering on the Iris dataset and visualize the clustering results.
9. **Data Collection Techniques**
   - Learn data collection through web scraping, APIs, and data connectors from suitable online sources.
10. **Association Rule Mining**
    - Perform association analysis on a given dataset and evaluate the generated rules and their accuracy.
11. **Recommendation System**
    - Build a recommendation system using a suitable dataset and evaluate its performance.
12. **Time Series Analysis**
    - Develop a time-series forecasting model and evaluate its prediction accuracy.
13. **Cartographic Visualization**
    - Create geographical and cartographic visualizations using datasets involving countries, Indian states, and districts.
14. **Text Mining and Visualization**
    - Perform text mining on a collection of documents and visualize significant terms using techniques such as Word Clouds.

Note

- The instructor may add, delete, modify, or tune experiments as required.
- Open-source tools such as **R**, **Python**, and related analytics libraries are recommended for conducting laboratory experiments.

DESIGN AND ANALYSIS OF ALGORITHMS LAB (BCS553)

Detailed Syllabus

1. Write programs for Recursive Linear Search and Binary Search.
2. Implement Heap Sort.
3. Implement Merge Sort.
4. Implement Selection Sort.
5. Implement Insertion Sort.
6. Implement Quick Sort.
7. Solve the Knapsack Problem using Greedy Method.
8. Implement Travelling Salesman Problem (TSP).
9. Find Minimum Spanning Tree using Kruskal's Algorithm.
10. Implement N-Queen Problem using Backtracking.
11. Analyze Quick Sort by computing time complexity for different input sizes and plot graph of execution time versus number of elements.
12. Analyze Merge Sort by computing time complexity for different input sizes and plot graph of execution time versus number of elements.
13. Implement 0/1 Knapsack Problem using:
    - Dynamic Programming Method
    - Greedy Method
14. Find Shortest Path using Dijkstra's Algorithm.
15. Find Minimum Cost Spanning Tree using Kruskal's Algorithm with Union-Find Method.
16. Find Minimum Cost Spanning Tree using Prim's Algorithm.
17. Implement:
    - All-Pairs Shortest Path using Floyd's Algorithm
    - Travelling Salesman Problem using Dynamic Programming
18. Design and implement Sum of Subsets Problem using Backtracking.
19. Find all Hamiltonian Cycles in a Connected Undirected Graph using Backtracking.

Note

The instructor may add, delete or modify experiments as required. Open-source tools such as C and C++ are recommended for conducting the lab sessions.

SOFTWARE ENGINEERING (BCS601)

Unit I: Introduction to Software Engineering

Introduction to Software Engineering, Software Components, Software Characteristics, Software Crisis, Software Engineering Processes, Comparison between Software Engineering and Conventional Engineering Processes, Software Quality Attributes.Software Development Life Cycle (SDLC) Models:
Waterfall Model, Prototype Model, Spiral Model, Evolutionary Development Model and Iterative Enhancement Model.

Unit II: Software Requirement Specification and Quality Assurance

Software Requirement Specification (SRS): Requirement Engineering Process, Requirement Elicitation, Analysis, Documentation, Review and Management of User Requirements.
Feasibility Study, Information Modeling, Data Flow Diagrams (DFD), Entity Relationship Diagrams (ERD), Decision Tables, SRS Documentation and IEEE Standards for SRS.
Software Quality Assurance (SQA): Verification and Validation, SQA Plans, Software Quality Frameworks, ISO 9000 Models and SEI-CMM Model.

Unit III: Software Design and Software Metrics

Basic Concepts of Software Design, Architectural Design and Low-Level Design.
Modularization, Structure Charts, Pseudo Code, Flowcharts, Coupling and Cohesion.
Design Strategies:
Function-Oriented Design, Object-Oriented Design, Top-Down Design and Bottom-Up Design.
Software Measurement and Metrics:
Halstead's Software Science, Function Point (FP) Metrics, Cyclomatic Complexity and Control Flow Graphs.

Unit IV: Software Testing

Software Testing Objectives, Unit Testing, Integration Testing, Acceptance Testing and Regression Testing.
Testing for Functionality and Performance, Top-Down and Bottom-Up Testing Strategies, Test Drivers and Test Stubs.
White Box Testing and Black Box Testing, Test Data Preparation, Alpha Testing and Beta Testing.
Static Testing Techniques:
Formal Technical Reviews, Peer Reviews, Walkthroughs, Code Inspection and Compliance with Design and Coding Standards.

Unit V: Software Maintenance and Project Management

Software Maintenance: Software Evolution, Need for Maintenance, Types of Maintenance – Preventive, Corrective and Perfective Maintenance, Cost of Maintenance.
Software Re-engineering and Reverse Engineering.
Software Configuration Management: Configuration Management Activities, Change Control Process, Software Version Control and CASE Tools.
Software Project Management:
Cost Estimation, Effort Estimation, Scheduling and Duration Estimation, COCOMO Model, Resource Allocation Models, Software Risk Analysis and Risk Management.

Text books:

- RS Pressman, Software Engineering: A Practitioners Approach, McGraw Hill.
- Pankaj Jalote, Software Engineering, Wiley
- Rajib Mall, Fundamentals of Software Engineering, PHI Publication.
- KK Aggarwal and Yogesh Singh, Software Engineering, New Age International Publishers.
- Ghezzi, M. Jarayeri, D. Manodrioli, Fundamentals of Software Engineering, PHI Publication.
- Ian Sommerville, Software Engineering, Addison Wesley.
- Kassem Saleh, "Software Engineering", Cengage Learning.
- P fleeger, Software Engineering, Macmillan Publication

BCDS601: BIG DATA AND ANALYTICS

Unit I: Introduction to Big Data

Introduction to Big Data, types of digital data, history and evolution of Big Data, Big Data platforms, drivers of Big Data, architecture and characteristics of Big Data, and the 5Vs of Big Data.
Big Data technology components, importance and applications of Big Data, security, compliance, auditing, data protection, privacy, and ethics in Big Data environments.
Big Data Analytics, challenges of conventional systems, intelligent data analysis, nature of data, analytical processes and tools, analysis versus reporting, and modern data analytics tools.

Unit II: Hadoop and MapReduce

History and evolution of Hadoop, Apache Hadoop, Hadoop Distributed File System (HDFS), Hadoop components, data formats, data analysis using Hadoop, scalability, Hadoop Streaming, Hadoop Pipes, and Hadoop Ecosystem.
MapReduce Framework: Concepts and architecture of MapReduce, working of MapReduce, development of MapReduce applications, unit testing with MRUnit, local testing, anatomy of a MapReduce job, handling failures, job scheduling, shuffle and sort process, task execution, MapReduce types, input and output formats, features, and real-world applications.

Unit III: Hadoop Distributed File System (HDFS)

Design and architecture of HDFS, concepts, advantages and challenges, file sizes, block abstraction, block sizes, data replication, file storage, reading and writing mechanisms in HDFS.
Java interfaces to HDFS, command-line interface, Hadoop file system interfaces, data flow, data ingestion using Flume and Sqoop, Hadoop archives.
Hadoop I/O concepts including compression, serialization, Avro, and file-based data structures.
Hadoop Environment: Cluster setup and installation, cluster specifications, Hadoop configuration, Hadoop security, administration, HDFS monitoring and maintenance, benchmarking, and Hadoop deployment in cloud environments.

Unit IV: Hadoop Ecosystem, YARN and NoSQL

Hadoop ecosystem components, scheduling mechanisms, Fair Scheduler and Capacity Scheduler.
Hadoop 2.0 features including NameNode High Availability, HDFS Federation, MRv2, YARN, and execution of MRv1 applications on YARN.
NoSQL Databases: Introduction and concepts of NoSQL databases.
MongoDB: Introduction, data types, document creation, updating and deletion, querying, indexing, and capped collections.
Apache Spark: Installation, Spark applications, jobs, stages, tasks, Resilient Distributed Datasets (RDDs), Spark job execution, and Spark on YARN.
Scala Programming: Introduction to Scala, classes and objects, data types and operators, control structures, functions, closures, and inheritance.

Unit V: Hadoop Ecosystem Frameworks

Applications of Big Data using Pig, Hive, and HBase.
Pig: Introduction, execution modes, comparison with databases, Grunt shell, Pig Latin language, user-defined functions, and data processing operators.

Hive: Architecture and installation, Hive Shell, Hive services, Hive Metastore, comparison with traditional databases, HiveQL, table creation, querying, sorting, aggregation, MapReduce scripts, joins, and subqueries.
HBase: Concepts, clients, practical examples, comparison with RDBMS, advanced usage, schema design, advanced indexing, and integration with distributed systems.
ZooKeeper: Cluster monitoring, coordination services, and application development using ZooKeeper. IBM Big Data Strategy, introduction to InfoSphere, BigInsights, BigSheets, and Big SQL.

Text Books and References

1. Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, Wiley.
2. *Big Data Black Book*, DT Editorial Services, Wiley.
3. Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, and Tom Deutsch, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill.
4. Thomas Erl, Wajid Khattak, and Paul Buhler, *Big Data Fundamentals: Concepts, Drivers and Techniques*, Prentice Hall.
5. Raj Kamal and Preeti Saxena, *Big Data Analytics*, McGraw-Hill Education.
6. Bart Baesens, *Analytics in a Big Data World: The Essential Guide to Data Science and its Applications (Wiley Big Data Series)*, John Wiley & Sons.
7. Arshdeep Bahga and Vijay Madisetti, *Big Data Science & Analytics: A Hands-On Approach*, VPT.
8. Anil Maheshwari, *Big Data* (Second Edition), McGraw-Hill.
9. Anand Rajaraman and Jeffrey David Ullman, *Mining of Massive Datasets*, Cambridge University Press.
10. Tom White, *Hadoop: The Definitive Guide*, O'Reilly Media.
11. Eric Sammer, *Hadoop Operations*, O'Reilly Media.
12. Chuck Lam, *Hadoop in Action*, Manning Publications.
13. Deepak Vohra, *Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools*, Apress.
14. Edward Capriolo, Dean Wampler, and Jason Rutherglen, *Programming Hive*, O'Reilly Media.
15. Lars George, *HBase: The Definitive Guide*, O'Reilly Media.
16. Alan Gates, *Programming Pig*, O'Reilly Media.
17. Michael Berthold and David J. Hand, *Intelligent Data Analysis*, Springer.
18. Bill Franks, *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, John Wiley & Sons.
19. Glenn J. Myatt, *Making Sense of Data*, John Wiley & Sons.
20. Pete Warden, *Big Data Glossary*, O'Reilly Media.

COMPUTER NETWORKS (BCS603)

Unit I: Introduction to Computer Networks

Introduction to Computer Networks, Goals and Applications of Networks, Categories of Networks, Organization of the Internet and Internet Service Providers (ISP).
Network Architecture: Layering Principles, Services, Protocols and Standards, OSI Reference Model, TCP/IP Protocol Suite.
Network Devices and Components.

Unit II: Physical Layer and Data Link Layer

Physical Layer:
Network Topology Design, Types of Connections, Transmission Media, Signal Transmission and Encoding, Network Performance, Transmission Impairments, Switching Techniques and Multiplexing.
Data Link Layer:
Framing, Error Detection and Correction, Flow Control Techniques, Elementary Data Link Protocols and Sliding Window Protocols.

Unit III: Medium Access Control and Network Layer

Medium Access Control (MAC) and Local Area Networks:
Channel Allocation, Multiple Access Protocols, LAN Standards, Link Layer Switches and Bridges, Learning Bridge and Spanning Tree Algorithms.
Network Layer:
Point-to-Point Networks, Logical Addressing, Internetworking Concepts, IP Addressing, CIDR, ARP, RARP, DHCP and ICMP.
Routing, Forwarding and Delivery, Static and Dynamic Routing, Routing Algorithms and Protocols, Congestion Control Algorithms and IPv6.

Unit IV: Transport Layer

Process-to-Process Delivery, Transport Layer Protocols: UDP and TCP.
Multiplexing, Connection Management, Flow Control and Retransmission, Window Management, TCP Congestion Control and Quality of Service (QoS).
Domain Name System (DNS), World Wide Web (WWW), Hyper Text Transfer Protocol (HTTP), Electronic Mail, File Transfer Protocol (FTP), Remote Login and Network Management.
Data Compression and Cryptography: Basic Concepts.

Text Books and References

1. Data Communications and Networking — Behrouz A. Forouzan, McGraw Hill.
2. Computer Networks — Andrew S. Tanenbaum, Prentice Hall.
3. Data and Computer Communications — William Stallings, Pearson Education.
4. Computer Networking: A Top-Down Approach — James F. Kurose and Keith W. Ross, Pearson.
5. Computer Networks: A Systems Approach — Larry Peterson and Bruce Davie, Morgan Kaufmann.
6. Understanding Communications and Networks — William A. Shay, Cengage Learning.
7. Computer Networks and Internets — Douglas Comer, Pearson Education.
8. TCP/IP Protocol Suite — Behrouz A. Forouzan, McGraw Hill.

Departmental Elective-III

CYBER FORENSIC ANALYTICS (BCAI061)

Unit I: Cyber Crime

Cyber Space, Cyber Crime, Criminal Behaviour, Jurisdictional Concerns, Jurisprudential Inconsistency, eCash Security, Prepaid Cards, Stored Value Cards, Mobile Payments, Internet Payment Services, Cyber Stalking, Cyber Extortion, Cyber Terrorism, Cyber Warfare, Cyber Weapons, ATM Frauds, Phreaking, and Internet Gambling.

Practical Component:

Key Logger

Email Fraud

Unit II: Cyber Forensics

Digital Devices, Hard Disk, Disk Characteristics, Disk Imaging, Data Carving, Techniques, Commercial Piracy, Soft Lifting, Steganography, Network Components, Port Scans, Wireshark, PCAP Analysis, Trojans and Backdoors, Botnets, DoS Attacks, DDoS Attacks, Honey Pots, Malware, Viruses, and Worms.

Practical Component:PCAP File Analysis – Case Study, Network Port Scan – Forensics

Unit III: Cyber Investigation

Concepts of Investigation, Cyber Investigation, Network Investigation, Investigating Audit Logs, Investigating Web Attacks, Investigating Computer Intrusions, Profiling, Cyber Criminal Profiling, Stylometric Techniques, Warranted Searches, Warrantless Searches, and Undercover Techniques.

Practical Component: Investigating Audit Logs, Investigating Web Attacks

Unit IV: Evidence Management

Evidence, Digital Evidence, Types of Evidence, Physical Evidence, Real Evidence, Circumstantial Evidence, Network Evidence, Evidence Collection, Evidence Analysis, Contextual Information, Evidence Management, Pre-Search Activities, On-Scene Activities, and Report Preparation.

Practical Component: Digital Evidence Analysis, Network Analysis

Unit V: Cyber Laws and Authorities

Information Technology Act 2000, Digital Signature, Electronic Governance, Secure Electronic Records, Regulation of Certifying Authorities, CERT-In, Electronic Signature Certificates, Penalties, Compensation, Future Trends, and Emerging Concerns.

Practical Component: Digital Signature, Text Books

Text Books:

1. Marjie T. Britz, "Computer Forensics and Cyber Crime", Pearson, 2013.

2. Garima Tiwari, "Understanding Laws– Cyber Laws And Cyber Crimes", Lexis Nexis, 2014.

3. Chuck Easttom, Jeff Taylor, "Computer Crime, Investigation, and the Law", Course Technology, 2018.

4. Eoghan Casey, "Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet", Eoghan Casey, 2018.

Departmental Elective-III

IMAGE ANALYTICS (BCDS061)

Unit I: Fundamentals of Digital Image Processing

Introduction to digital image processing, image processing systems, image acquisition, sampling and quantization, pixel relationships, and mathematical tools used in image processing. Intensity transformation functions including image negatives, logarithmic transformations, and power-law transformations. Histogram processing, color fundamentals, spatial filtering, smoothing filters, and sharpening filters.

Practical Components (Python/MATLAB):

1. Apply various intensity transformation functions.
2. Compute and plot image histograms using image processing toolboxes.
3. Implement color image smoothing and sharpening techniques.

Unit II: Morphological Image Processing

Fundamentals of morphological image processing, erosion and dilation operations, opening and closing operations, hit-or-miss transform, basic morphological algorithms, morphological reconstruction, and grayscale morphology.

Practical Components (Python/MATLAB):

1. Implement morphological operations.
2. Implement morphological reconstruction.
3. Implement grayscale morphology.

Unit III: Colour Image Processing and Segmentation

Colour fundamentals, colour models, pseudo-colour image processing, full-colour image processing, colour transformations, image smoothing and sharpening. Colour-based image segmentation, active contours using snakes and level sets, noise in colour images, and colour image compression. Thresholding techniques including basic global thresholding, Otsu's optimum global thresholding method, multiple thresholds, variable thresholding, region growing, and region splitting and merging methods.

Practical Components (Python/MATLAB):

1. Implement Otsu's optimum global thresholding method.
2. Implement image smoothing and sharpening.
3. Implement image segmentation using active contours (Snakes or Level Sets).

Unit IV: Feature Extraction

Feature representation and boundary preprocessing. Boundary feature descriptors including basic boundary descriptors, shape numbers, Fourier descriptors, and statistical moments. Regional feature descriptors such as topological descriptors, texture descriptors, and moment invariants. Principal component analysis as a feature descriptor and whole-image features. Scale-Invariant Feature Transform (SIFT).

Practical Components (Python/MATLAB):

1. Implement boundary feature descriptors.
2. Implement topological and texture descriptors.
3. Implement Scale-Invariant Feature Transform (SIFT).

Unit V: Image Pattern Classification

Introduction to patterns and pattern classes. Pattern classification using prototype matching, minimum-distance classifier, correlation-based 2D prototype matching, SIFT feature matching, and structural prototype matching. Optimum (Bayes) statistical classifiers.
Neural networks and deep learning concepts including perceptrons, multilayer feedforward neural networks, and deep convolutional neural networks (CNNs).

Practical Components (Python/MATLAB):

1. Implement minimum-distance classification algorithm.
2. Implement Bayes statistical classification algorithm.
3. Implement deep convolutional neural networks (CNNs).

Text Books:

1. Rafael C Gonzalez, Richard E Woods, "Digital Image Processing", 4th Edition, Pearson, 2018.
2. Kenneth R. Castleman, Digital Image Processing Pearson, 2006.
3. Anil K.Jain, "Fundamentals of Digital Image Processing", Person Education, 2003.

Departmental Elective-III

SOCIAL MEDIA ANALYTICS AND DATA ANALYSIS (BCAM061)

Unit I: Introduction to Social Media

Introduction to social media, social media landscape, social media analytics (SMA), and its importance. Application of SMA in small and large organizations and across different social media platforms. Types of social networks including friendship networks, user-generated content networks, and affiliation networks. Concepts of sociograms and sociometric studies.

Basics of social media business models, web search engines, digital advertising, and applications of social media analytics in various domains.

Unit II: Graphs and Matrices

Introduction to graph theory concepts including adjacency matrices, paths, connectivity, distance measures, and Breadth-First Search (BFS). Overview of network datasets.

Study of nodes, ties, influencers, making connections, and link analysis. Random graphs, network evolution, weighted networks, and hypergraphs.

Unit III: Network Fundamentals

Network structures, equivalence, homophily, clustering, snowball sampling, contact tracing, and random walks.

Ego-centered networks, dominance hierarchies, third-party records, affiliation networks, citation networks, peer-to-peer networks, recommender networks, and biological networks.

Unit IV: Social Network Modeling and Analytics

Social contexts including affiliation, identity, social capital, structural holes, and structural balance. Predictive modeling, descriptive modeling, community detection, and anomaly detection techniques.

Facebook Analytics: Introduction, parameters, demographics, page audience analysis, reach analysis, and engagement analysis.

Google Analytics: Implementation technology, limitations, performance concerns, privacy issues, working principles, and Google Website Optimizer.

Unit V: Processing, Visualizing, and Analyzing Social Media Data

Data processing and visualization techniques. Influence maximization, link prediction, collective classification, and applications in advertising and game analytics. Methods for collecting, visualizing, and exploring social media data.

Social network and web data analytics methods including clickstream analysis, A/B testing, online surveys, web crawling, and indexing.

Introduction to Natural Language Processing (NLP) techniques for micro-text analysis, trend analysis, social influencers, opinion spread, and judgment analysis.

Text and Reference Books:

1. Matthew Ganis, Avinash Kohirkar, Social Media Analytics: Techniques and Insights for Extracting Business Value Out of Social Media Pearson 2016
2. Jim Sterne, Social Media Metrics: How to Measure and Optimize Your Marketing Investment Wiley Latest edition
3. Brian Clifton, Advanced Web Metrics with Google Analytics, John Wiley & Sons; 3rd Edition edition (30 Mar 2012)
4. Ganis/Kohirka, SOCIAL MEDIA ANALYTICS Paperback – 29 September 2016 by Pearson.

Departmental Elective-III

CDS062: MACHINE LEARNING TECHNIQUES

Unit I Introduction to Machine Learning: Learning, types of learning, well-defined learning problems, designing a learning system, history of machine learning. Introduction to machine learning approaches including Artificial Neural Networks, Clustering, Reinforcement Learning, Decision Tree Learning, Bayesian Networks, Support Vector Machines, and Genetic Algorithms. Issues in Machine Learning. Data Science versus Machine Learning.

Unit II

Regression: Linear Regression and Logistic Regression.
Bayesian Learning: Bayes theorem, concept learning, Bayes optimal classifier, Naïve Bayes classifier, Bayesian belief networks, Expectation-Maximization (EM) algorithm.
Support Vector Machines: Introduction, support vector kernels (Linear, Polynomial, and Gaussian kernels), hyperplanes and decision surfaces, properties of SVM, and issues in SVM.

Unit III

Decision Tree Learning: Decision tree learning algorithm, inductive bias, inductive inference with decision trees, entropy and information theory, information gain, ID3 algorithm, and issues in decision tree learning.
Instance-Based Learning: k-Nearest Neighbour (k-NN) learning, locally weighted regression, radial basis function networks, and case-based learning.

Unit IV

Artificial Neural Networks: Perceptrons, multilayer perceptrons, gradient descent and delta rule, multilayer networks, derivation of backpropagation algorithm, generalization, unsupervised learning using Self-Organizing Maps (SOM) and its variants.
Deep Learning: Introduction to deep learning, concept of Convolutional Neural Networks (CNNs), convolutional layers, activation functions, pooling layers, fully connected layers, one-dimensional and two-dimensional convolution, training of deep networks, and applications such as diabetic retinopathy detection, smart speakers, and self-driving cars.

Unit V

Reinforcement Learning: Introduction to reinforcement learning, learning tasks, practical applications, Markov Decision Process (MDP), Q-Learning, Q-Learning algorithm, applications of reinforcement learning, and introduction to Deep Q-Learning.
Genetic Algorithms: Introduction, components of genetic algorithms, reproduction cycle, crossover, mutation, genetic programming, models of evolution and learning, and applications of genetic algorithms.

Text books:

1. Tom M. Mitchell, —Machine Learning, McGraw-Hill Education (India) Private Limited, 2013.
2. Ethem Alpaydin, —Introduction to Machine Learning (Adaptive Computation and Machine Learning), The MIT Press 2004.
3. Stephen Marsland, —Machine Learning: An Algorithmic Perspective, CRC Press, 2009.
4. Bishop, C., Pattern Recognition and Machine Learning. Berlin: Springer-Verlag.

SOFTWARE ENGINEERING LAB (BCS651)

Detailed Syllabus

For a given case study or problem statement, perform the following experiments:

1. Prepare a Software Requirement Specification (SRS) document based on IEEE standards.
2. Draw the Use Case Diagram and identify the role of each actor, along with preconditions, postconditions and functions of each use case.
3. Draw the Activity Diagram.
4. Identify weak and strong classes and draw the Class Diagram.
5. Draw Sequence Diagrams for any two scenarios.
6. Draw the Collaboration Diagram.
7. Draw the State Chart Diagram.
8. Draw the Component Diagram.
9. Perform Forward Engineering in Java (Model to Code Conversion).
10. Perform Reverse Engineering in Java (Code to Model Conversion).
11. Draw the Deployment Diagram.

Note

The instructor may add, delete or modify experiments as required. Open-source tools such as OpenOffice, LibreOffice, JUnit, OpenProj, GanttProject, dotProject, ArgoUML and StarUML are recommended for conducting the lab sessions.

BCDS651: BIG DATA AND ANALYTICS LAB

1. **Hadoop Installation and Configuration**
   - Download and install Hadoop.
   - Study different Hadoop operating modes.
   - Understand startup scripts and Hadoop configuration files.
2. **HDFS File Management Operations**
   - Perform file and directory management tasks in Hadoop:
     - Adding files and directories
     - Retrieving files
     - Deleting files
   - Understand data movement and storage workflows in HDFS.
3. **Matrix Multiplication Using MapReduce**
   - Implement matrix multiplication using the Hadoop MapReduce framework.
4. **Weather Data Analysis Using MapReduce**
   - Develop a MapReduce program to process and analyze weather sensor data collected from multiple locations.
5. **Word Count Using MapReduce**
   - Implement the Word Count application to understand the MapReduce programming paradigm.
6. **K-Means Clustering Using MapReduce**
   - Implement K-Means clustering algorithm using Hadoop MapReduce for large-scale data processing.
7. **Hive Installation and Practice**
   - Install Apache Hive and perform practical exercises using HiveQL.
8. **HBase and Thrift Installation**
   - Install Apache HBase and Thrift.
   - Perform practical examples demonstrating data storage and retrieval.
9. **Database Connectivity**
   - Practice importing and exporting data between Hadoop and various database systems.
10. **Pig Latin Scripting**
    - Write Pig Latin scripts for:
      - Sorting
      - Grouping
      - Joining
      - Projection
      - Filtering operations
11. **Word Count Using Pig**
    - Execute Pig Latin scripts to perform Word Count analysis.
12. **Temperature Analysis Using Pig**
    - Develop Pig Latin scripts to determine the maximum temperature recorded for each year from a weather dataset.

Note

- The instructor may add, delete, modify, or tune experiments as required.
- Open-source Big Data tools such as **Hadoop, HDFS, Hive, Pig, HBase, MapReduce, Spark, and NoSQL databases** are recommended for conducting laboratory experiments.

COMPUTER NETWORKS LAB (BCS653)

Detailed Syllabus

1. Implement Stop-and-Wait Protocol and Sliding Window Protocol.
2. Study Socket Programming and Client-Server Model.
3. Write programs to simulate ARP and RARP protocols.
4. Write programs to simulate PING and TRACEROUTE commands.
5. Create HTTP Socket programs for web page upload and download.
6. Implement Remote Procedure Call (RPC).
7. Perform Subnetting and IP Address Calculations.
8. Develop TCP Socket Applications:
   - Echo Client and Echo Server
   - Chat Application
   - File Transfer Application
9. Develop TCP and UDP Socket Applications:
   - DNS
   - SNMP
   - File Transfer
10. Study Network Simulator (NS) and simulate Congestion Control Algorithms.
11. Perform case studies on Routing Algorithms:
    - Link State Routing
    - Flooding
    - Distance Vector Routing
12. Learn handling and configuration of networking hardware such as RJ-45 connectors, CAT-6 cables and crimping tools.
13. Configure networking devices such as Routers, Hubs and Switches using real devices or simulators.
14. Practice networking commands and services:
    - ping
    - traceroute
    - nslookup
    - arp
    - telnet
    - ftp
15. Perform Network Packet Analysis using tools like Wireshark and tcpdump.
16. Perform Network Simulation using Cisco Packet Tracer, NetSim, OMNeT++, NS2 and NS3.
17. Implement Socket Programming using TCP and UDP:
    - DNS Application
    - Date and Time Client/Server
    - Echo Client/Server
    - Iterative and Concurrent Servers

Note

The instructor may add, delete or modify experiments as required. Open-source tools such as C, C++, Java, NS3, Mininet, OPNET, TCPDump and Wireshark are recommended for conducting the lab sessions.

Semester IV

Open Elective- 01

IDEA TO BUSINESS MODEL (MOE-060)

Course Objectives

1. To motivate students to understand the fundamentals of establishing and sustaining a business enterprise.
2. To provide students with knowledge about identifying, evaluating, and selecting innovative business ideas and business models.
3. To develop entrepreneurial competencies through entrepreneurship development, production planning and control, and communication systems for effective business management.

Unit I – Introduction

Search for a business idea. Methods of choosing a business idea. Product idea and product selection. Product adoption process. Product innovation. Production planning and development strategy. Development of new product ideas.

Unit II – Introduction to Entrepreneurship

Meaning and concept of entrepreneurship. Difference between entrepreneurship and wage employment. Functions of an entrepreneur. Entrepreneur vs. manager. Role of entrepreneurship in economic development. Barriers to entrepreneurship.

Unit III – Entrepreneurial Development

Types of entrepreneurs. Competencies required for entrepreneurship. Creative and design thinking. Entrepreneurial decision-making process. Entrepreneurship Development Programme (EDP). Evaluation of EDP. Entrepreneurial development training.

Unit IV – Production Planning and Control

Production system and design of production systems. Types of production systems. Production Planning and Control (PPC). Steps involved in PPC.

Unit V – Communication and Negotiation Skills

Importance of communication systems. Barriers to communication. Effective listening skills. Power of communication and presentation. Personal selling. Risk-taking and resilience. Negotiation skills.

Text Books

1. Entrepreneurship Development – Sangeeta Sharma, Kindle Edition.
2. Production and Operations Management – Kanishka Bedi.
3. Marketing Management – Philip Kotler.
4. The Business Model Book: Design, Build and Adapt Business Ideas that Drive Business Growth – Adam Bock and Gerard George.

Software Project Management (MOE-068)

Unit I: Project Evaluation and Project Planning

Introduction to Software Project Management, Importance of Software Project Management and Project Management Activities.
Software Project Methodologies and Categorization of Software Projects.
Project Planning:
Setting Objectives, Management Principles, Management Control and Project Portfolio Management.
Cost-Benefit Evaluation, Risk Evaluation, Strategic Program Management and Stepwise Project Planning.

Unit II: Project Life Cycle and Effort Estimation

Software Process and Process Models, Selection of Process Models.
Rapid Application Development (RAD), Agile Methods, Dynamic System Development Method (DSDM) and Extreme Programming (XP).
Managing Iterative Processes.
Software Estimation:
Basics of Software Estimation, Effort Estimation and Cost Estimation Techniques.
COSMIC Full Function Points, COCOMO II and Parametric Productivity Models.

Unit III: Activity Planning and Risk Management

Objectives of Activity Planning, Project Scheduling, Activity Sequencing and Scheduling.
Network Planning Models:
Formulation of Network Models, Forward Pass and Backward Pass Techniques, Critical Path Method (CPM) and PERT Technique.
Risk Management:
Risk Identification, Risk Assessment, Risk Planning and Risk Management.
Monte Carlo Simulation, Resource Allocation, Creation of Critical Paths and Cost Scheduling.

Unit IV: Project Management and Control

Framework for Project Management and Control.
Data Collection and Progress Visualization.
Cost Monitoring and Earned Value Analysis (EVA).
Project Tracking, Prioritizing Monitoring and Change Control.
Software Configuration Management (SCM), Contract Management and Managing Contracts.

Unit V: Staffing in Software Projects

Managing People and Organizational Behavior.
Staff Selection Methods, Motivation Techniques and the Oldham-Hackman Job Characteristic Model.
Stress Management, Health and Safety Issues, Ethical and Professional Concerns.
Team Work and Decision Making.
Organizational Structures, Distributed and Virtual Teams, Communication Plans and Leadership Skills.

Text Books

1. Software Project Management — Bob Hughes, Mike Cotterell and Rajib Mall, Fifth Edition, McGraw Hill, 2012.
2. Effective Software Project Management — Robert K. Wysocki, Wiley Publication, 2011.
3. Software Project Management — Walker Royce, Addison-Wesley, 1998.
4. Managing Global Software Projects — Gopalaswamy Ramesh, McGraw Hill Education (India), 2013.