Semester IV

CBCS Course Outline M. Sc (IT) Semester-IV

Semester-IV (26 Credit Semester)						
Course Code	**Course Name**	**Paper category**	**Hours / Week**			**Credits**
			L	T	P	
20 Core Credit Units						
MIT-4T1-C	Data Warehousing and Mining	Core	3	0	4	5
MIT-4T2-C	Theory of Computation & Formal Languages	Core	4	0	4	5
MIT-4T3-C	Major Project Work	Core				10
6 Elective Credit Units						
MIT-4E1-DCE	Bio Informatics	DCE	3	0	0	3
MIT-4E2-DCE	Cloud and Grid Computing	DCE	3	0	0	3
MIT-4E3-DCE	Information Security and Networks	DCE	3	0	0	3
MIT-4E4-DCE	Pattern Recognition	DCE	3	0	0	3
MIT-4E5-DCE	Research Methodologies	DCE	3	0	0	3

Course Title: Data Warehousing & Data Mining
Course Code: MIT-4T1-C

UNIT I

Overview of decision support systems: Organizational need for strategic information, Failures of past decision-support systems, operational versus decision-support systems, data warehousing – the only viable solution, data warehouse defined. Data warehouse – The building Blocks: Defining Features, data warehouses & data marts, overview of the components, metadata in the data warehouse
Defining the business requirements: Dimensional analysis, information packages – a new concept, requirements gathering methods, requirements definition: scope & content

UNIT II

Principles of dimensional modeling: Objectives, From Requirements to data design, the STAR schema, STAR Schema Keys, Advantages of the STAR Schema
Dimensional Modeling: Updates to the Dimension tables, miscellaneous dimensions, the snowflake schema, aggregate fact tables.

UNIT III

OLAP in the Data Warehouse: Demand for Online analytical processing, need for multidimensional analysis, fast access & powerful calculations, limitations of other analysis methods, OLAP is the answer, OLAP definitions & rules, OLAP characteristics, major features & functions, general features, dimensional analysis, what are hypercubes? Drill-down & roll-up, slice-&-dice or rotation, OLAP models, overview of variations, the MOLAP model, the ROLAP model, ROLAP versus MOLAP, OLAP implementation considerations

UNIT IV

Data Mining Basics: What is Data Mining, Data Mining Defined, The knowledge discovery process, OLAP versus data mining, data mining & the data warehouse, Applications of data Mining ,Major Data Mining Techniques, Cluster detection, decision trees, memory-based reasoning, link analysis, neural networks, genetic algorithms.

REFERENCE BOOKS:

1. Paul Raj Poonia, "Fundamentals of Data Warehousing", **John Wiley & Sons**
2. Sam Anahony, "Data Warehousing in the real world: A practical guide for building decision support systems", **John Wiley**
3. Alex Berson, Stephen J. Smith, "Data Warehousing, Data Mining & OLAP", **Tata McGraw Hill**
4. W. H. Inmon, "Building the operational data store", **John Wiley**
5. Kamber& Han, "Data Mining Concepts & Techniques", **Hartcourt India P. Ltd.**

Course Title: Theory of Computation & Formal Languages
Course Code: MIT-4T2-C

UNIT I

Introduction to Set Theory, Introduction to Automata, Alphabets, String, Languages, Strings, Kleen Star, Kleen Plus, Introduction to Finite automata, Deterministic finite automata, DFA notations (Transition Table, Transition Table) Languages Accepted by DFA, Designing DFA by pattern recognition, Application of DFA, Non-deterministic Finite Automata, Language accepted by NDFA, Conversion of DFA to NDFA, Overview of e-NDFA. DFA vs NDFA vs e-NDFA,

UNIT II

Finite Automata & Regular Expressions: Definition, Basic Regular Expressions, Obtaining regular expression: Using Basics Expressions & State Elimination Method, Application of regular Expressions, Regular Languages: Definition, Properties, and Pumping Lemma for Regular Languages, Decidable and Closure properties of Regular languages. Limitation of Finite Automata, Equivalence & minimization of DFA.

UNIT III

Introduction to Grammar, Chomsky Hierarchy, Generation of Grammar from Finite Automata & Regular Expressions, Derivation Trees, Left most and Right Most Derivation Trees, Ambiguous Grammar, Context Free Grammar and its applications, properties of context free languages. Introduction to Push down Automata, Transition & Graphical Representation.

UNIT IV

Introduction to Linear bounded Automata & Turing Machine. Turing machine model, transition table Instantaneous definition, Acceptance and Construction of Turing machine. Standard Turing machine, Transducers, Lexical and Syntax Analyse: Parsing, Top down Parsing (Predictive parsers, LL (K), recursive Descent) and Bottom up Parsing (LR (K)).

REFERENCES:

1. Hopcroft, J., & Ullman, J., "Introduction to Automata Theory, Languages & Computation", **Pearson Education**
2. Hopcroft J, R. Motwani, & J. Ullman, "Introduction to Automata Theory, Languages & Computation", **Pearson Education**
3. P. Linz, "Introduction to Formal Languages & Automata", Jones &Barlett PWS Publishing Company
4. Donald Knuth, "The Art of Computer Programming", **Prentice Hall**

5. Fintite Automata and Formal Languages"- A Simple Approach, A.M Padma Ready Pearson Education

Course Title: Project Work
Course Code: MIT-4T3-C

Project Work

Course Title: Bioinformatics
Course Code: MIT-4E1-DCE

UNIT 1

Introduction to biology, Living cell: characteristics & functions, basics of molecular biology: DNA, RNA, gene, genetic code, processes of transcription, translation, splicing etc. ,basics of proteins, sequences, structure & functions of proteins, Introduction tobioinformatics: Definitions & concepts, & the role of bioinformatics, Types of biological data.

UNIT 2

Introduction to databases, types of Biological Databases – flat file databases, relational databases, object-oriented databases, introduction to biological databases**,** Sequence databases (EMBL, GenBank, DDBJ, SWISS-PROT, PIR, TrEMBL), search engines: SRS,ENTREZ. BLAST, FASTA.

UNIT 3

Basic overview of algorithms: Space & time complexity. Need & significance of sequence alignment. Global & local alignments , techniques of sequence alignments: Doplot, Needleman-wunsch algorithm for global alignment, Smith-waterman algorithm for local alignments ,substitution matrices: PAM, Blosum etc. Gene prediction(ab initio & similarity based). Ontologies in bioinformatics: need for ontologies.

UNIT 4

Languages in Bioinformatics: Bioperl, GEML, BSML, NeuroML&RasMolCrytal. Applications of Bioinformatics: Accessing Sequence Information using Bioperl: Bioperl Installation, Bioperl Modules, Object Oriented Programming,UsingBioperl& Write-Sequence()function.. Advanced applications of bioinformatics: Drug design, Applications in molecular medicine etc.

REFERENCE BOOKS:

1. Harshawardhan P. Bal.” Bioinformatics Principles & Applications” O’Reily
2. T. K. Attwood & D J Parry-Smith, “Introduction to bioinformatics”, **Pearson Education**
3. Jean-Michel Claveriw, CerdricNotredame, “Bioinformatics – A beginner’s Guide”, **WILEY DreamTech India Pvt**
4. S.C. Rastogi, N. Mendiratta” Bioinformatics Methods &Applications”**TMH**
5. Krane ,”Bioinformatic”, **Pearson Education**
6. B. Bergeron, “Bioinformatics Computing”, **Pearson**
7. D. E. Krene& M.L. Payma, “Fundamental concepts of Bioinformatics”, **Pearson**

Course Title: Cloud & Grid Computing
Course Code: MIT-4E2-DCE

UNIT I

Cloud Computing: Introduction to cloud & Cluster computing, cloud computing vs cluster computing, Evolution of cloud computing, principles of cloud computing, cloud Computing architecture, Cloud computing applications. Cloud service models (IaaS, PaaS, SaaS), could Deployment models (Public, Private, hybrid& community models). Challenge and Security Issues

UNIT II

Grid Computing: Introduction to Grid Computing, Characteristics, grid computing Architecture, grid complexity levels and topologies, grid components and grid layers, applications of Grid Computing. Grid security issues: Authorization and Authentication methods, Grid computing vs Cloud Computing Cluster Computing.

UNIT III

Introduction to OGSA, Services, Schema and architecture, Overview of OGSI, Virtualization: Virtual machines and visualization of clusters and data centres, levels of virtualization, virtualization structures and tools and mechanism, virtualization of CPU, Memory & I/O devices

REFERENCE BOOKS:

1. Mastering Cloud Computing, Paperback-1 Feb 2013 by Buyya, Vecchiola&Selvi.
2. Cloud Computing: Concepts, Technology & Architecture, 1e Paperback-2014 by Erl. .
3. Grid & Cloud Computing, a business perspective on Technology & Applications, Springer by **Stanoevska-Slabeva**, Katarina, **Wozniak**, Thomas, **Ristol**, Santi (Eds.)

Course Title: Information Security & Networks
Course Code: MIT-4E3-DCE

UNIT 1

Introduction: Security trends, the OSI Security Architecture, Security attacks, Security mechanisms, Security services, Model for Network Security. Encryption: Symmetric & Asymmetric Encryption, Symmetric Cipher model, Substitution techniques (Mono-alphabetic & Poly-alphabetic) & Transposition Technique, Rotor Machines, Block and Stream ciphers, Modern Round Ciphers: DES, AES, Introduction to Steganography. Encryption vs Steganography.

UNIT 2

Introduction to Number Theory: Prime Numbers, Fermat's & Euler's theorem & discrete logarithms. Public Key Cryptography & RSA. Key management: Diffie-Hellman Key Exchange, Elliptic Curve Arithmetic & Elliptic Curve Cryptography. Message Authentication & Hash functions. Digital Signatures & Authentication protocols. Network Security Applications: IP Security

UNIT 3

Mechanics of Routing Protocols: Routing , Static and Dynamic Routing, Distant vector routing & link state routing, Internet working with Dissimilar Protocols, Protocol design Consideration: Simplicity, flexibility, Optimality, Overhead & Scaling, Operation above Capacity, forward compatibility. Migration: Routing Algorithms & addressing parameters, making multi-protocol operation possible,

Network Security: Security Features in Wireless Adhoc & Wireless Sensor Networks.

REERENCES:

1. William Stalling ," Cryptography & Network Security", **Pearson Education**
2. Radia Perlman, "Interconnections: Bridges, Routers switches & Internet-working protocols **Pearson education**
3. Mark Sportack, "IP Routing Fundamentals", **Pearson Education**
4. Gerard J. Holzmann, "Design & Validation Computer Protocols", **Prentice Hall**

Course Title: Pattern Recognition
Course Code: MIT-4E4-DCE

UNIT I

Introduction to Pattern Recognition, Pattern recognition approaches, pattern recognition applications, Pattern recognition systems: Sensing, Segmentation & grouping, Feature extraction, Classification, Post processing. Design cycle, Statistical pattern recognition, syntactic (structural) pattern recognition, neural pattern recognition, Parameter estimation techniques, Non-parametric Pattern Recognition

UNIT II

Introduction to Bayesian decision theory, Bayesian estimation: Gaussian distribution, ML estimation, EM algorithm, Feature selection & extraction techniques, Linear Discriminant Functions, Introduction to Dimension Reduction techniques.

UNIT III

Supervised Learning, Unsupervised Learning, Reinforcement Learning, Support Vector Machines, Bayesian Belief Nets, Bayes estimation (learning). Density estimation, Probabilistic Classification. Maximum Likelihood parameter estimation. Bias & Biased Estimates.

REERENCES:

1. Machine Learning. Tom Mitchell. First Edition, **McGraw- Hill, 1997**.
2. Introduction to Machine Learning Edition 2, **by EthemAlpaydin**
3. Introduction to Artificial Intelligence – **RajendraAkerkar, PHI.**
4. Artificial Intelligence – A Modern Approach. Second Edition, Stuart Russel, Peter Norvig, **PHI/Pearson Education.**
5. Artificial Intelligence, 3rd Edition, Patrick Henry Winston., **Pearson Edition**
6. Artificial Intelligence & Expert Systems – **Patterson PHI**
7. Expert Systems: Principles & Programming- Fourth Edn, Giarrantana/ Riley, Thomson

Course Title: Research Methodology
Course Code: MIT-4E5-DCE

UNIT I

Concept of Research: Introduction to research, Meaning of Research, Objectives of Research, Motivation in Research, Types of Research, Research Approaches, Significance of Research, Research Methods versus Methodology, Research & Scientific Method, Importance of Knowing How Research is Done, Research Process, Criteria of Good Research. What is a Research Problem, Selecting the Problem, Necessity of Defining the Problem, Technique Involved in Defining a Problem.

UNIT II

Research Design: Meaning of Research Design, Need for Research Design, Features of a Good Design, Different Research Designs, Basic Principles of Experimental Designs, Sampling Design: Steps in Sampling Design, Criteria of Selecting a Sampling Procedure, Characteristics of a Good Sample Design, Different Types of Sample Designs. Methods of Data Collection.

UNIT III

Testing of Hypothesis: What is a Hypothesis, Basic Concepts Concerning Testing of Hypothesis, Procedure for Hypothesis Testing, Data Analysis: Data Preparation, Univariate analysis (frequency tables, bar charts, pie charts, percentages), Bivariate analysis – Cross tabulations & Chi-square test.

TEXT BOOKS:

1.Research Methodology – C.R.Kothari

REFERENCE BOOKS:

1.Business Research Methods – Donald Cooper & Pamela Schindler, TMGH, 9th edition
2.Business Research Methods – Alan Bryman& Emma Bell, Oxford University Press.