R V R & J C COLLEGE OF ENGINEERING *(Autonomous)*,
CHOWDAVARAM, GUNTUR-19
B.Tech., Computer Science & Engineering (Data Science)
(w.e.f. the academic year 2021-2022)
Scheme (R20)

Semester V (Third Year)

S. No.	Course Code	Course Title	Hours Per Week			Scheme of Examination			Category Code
			L	T	P	SES	EXT	Credits	
1	CD311	Automata and Compiler Design	2	1	-	30	70	3	PC
2	CD312	Data Communications and Networking	3	-	-	30	70	3	PC
3	CD313	Machine Learning	2	1	-	30	70	3	PC
4	CD314 (CDEL02)	Distributed and Cloud Computing (Professional Elective-I)	3	-	-	30	70	3	PE
5	CD315 (JOEL26)	Business Intelligence Tools (Open/Job Oriented Elective-I)	3	-	-	30	70	3	OE
6	CD351	Machine Learning Lab	-	-	2	30	70	1.5	PC
7	CD352	Distributed and Cloud Computing Lab (Professional Elective-I lab)	-	-	2	30	70	1.5	PC
8	CD353	Summer Internship	-	-	-	100	-	1.5	PR
9	CDSL3	Softskills (Skill Oriented Course-III)	1	-	2	100	-	2	SC
10	CDV	Smart Coding and Practicing							
		Total	14	2	6	410	490	21.5	

Category	Credits
Skill Course	2.0
Professional Core Course	13.5
Professional Elective	3.0
Open Elective	3.0
Total credits	21.5

CD311	Automata and Compiler Design	L	T	P	C
		2	1	0	3

Course Objectives:

The main objectives of this course are to:

1. Introduce the types of Finite Automata and properties of Regular Expressions.
2. Explain Context-Free Grammars and Push Down Automata
3. To discuss the phases of compiler.
4. To demonstrate parsing techniques and syntax direct translation schemes.
5. To teach the intermediate code forms and code generation.

Course Outcomes:

The students will able to:

1. Explain the fundamental concepts of Automata and Formal languages. .
2. Apply the knowledge of Automata Theory, Formal languages, Grammars & Regular Expressions for solving various problems. Design PDAs for various languages.
3. Demonstrate through knowledge on the phases of compiler and Implement Parsers.
4. Write different SDT schemes and Design code generator through optimized intermediate code forms and apply the various code optimization methods, and runtime allocation strategies.

Course Content:

(CO-1) 12 periods

UNIT – I

Finite Automata: Introduction, An Informal picture of finite automata, Deterministic finite automata (DFA) - Definition, notations and the language of DFA, Non deterministic finite automata (NFA) – Definition, notations and the language of NFA, Equivalence of DFA and NFA Finite Automata and Regular Expressions- Converting DFA to Regular Expressions, converting Regular Expressions to Automata, Closure Properties of Regular Languages.

Context Free Grammars: Context Free Grammars, Parse Trees, Constructing parse trees, derivations and parse trees, ambiguous grammars.

Context free languages: Normal form's for context- Free grammars, the pumping lemma for context free languages.

UNIT – II **(CO-2) 12 Periods**

Pushdown Automata: Definition of the Pushdown automata, the languages of PDA, Equivalences of PDA's and CFG's.

Introduction to Compiling: Compilers - Analysis of the source program - Phases of a compiler- Cousins of the Compiler - Grouping of Phases.

Lexical Analysis: Role of Lexical Analyzer - Input Buffering –The Lexical Analyzer Generator-LEX, Use of Lex, structure of LEX program.

UNIT – III **(CO-3) 12 Periods**

Syntax Analysis: Introduction: Role of the parser - Top-Down parsing, Predictive parsing, LL (1) Parser.

Bottom-up parsing - Shift Reduce Parsing, LR Parsers - SLR Parser, Canonical LR Parser, and LALR Parser- The Parser Generator-YACC.

UNIT – IV **(CO-4) 10 Periods**

Syntax Directed Translation: Syntax Directed definition- construction of syntax trees, Bottom-up evaluation of S-attribute Definitions-L-attribute Definitions, Intermediate Code Generator – Register and address descriptors.

Code Generation: Issues in the design of code generator, a simple Code generator.

Code Optimization: Introduction, Peephole Optimization.

Learning Resources:

Text Books:

1. Introduction to Automata Theory, Languages, and Computation John E. Hopcroft, Rajeev Motwani and Jeffrey D. Ullman.
2. Compilers: Principles, Techniques and Tools, V. Aho, R. Sethi and J. Ullman.

Reference Books:

1. Elements of the Theory of Computation, Harry R. Lewis and Christos H.Papadimitriou.
2. Automata and Computability, Dexter C.Kozen.
3. Alfred V.Aho, Jeffrey D. Ullman, Principles of Compiler Design, Narosa publishing,2002.
4. Lex&Yacc - John R. Levine, Tony Mason, Doug Brown, 2nd Edition, O'reilly.
5. Engineering a Compiler - Keith Cooper & Linda Torezon, 2nd Edition Elsevier.

CD312	Data Communications and Networking	L	T	P	C
		3	0	0	3

Course Objectives:

1. To understand the concept of data communication, data link layer design issues, and protocol architecture of OSI & TCP/IP model and process between computer networks and switching components in telecommunication systems.
2. To understand the features of various protocol layers.
3. To understand various routing and congestion control algorithms.
4. To understand the transport and application layer protocols.

Course Outcomes:

After successful completion of the course student shall be able to:

1. Describe the basics of data communication and computer networks.
2. Illustrate the working of Data link layer in computer networks.
3. Compare protocols used in Network layer of the computer network models for communication.
4. Analyze the services and features of Transport and Application layers in networks.

Course Content:

UNIT – I **(CO1) (13 Periods)**

Introduction: Data communications, Uses of Computer networks, Networks, the internet, protocols and standards, layered tasks, the OSI model, layers in the OSI model, TCP/IP Protocol suite, addressing, Physical Layer: Transmission Media-Guided and Unguided, Multiplexing- FDM, WDM, TDM, Switching.

UNIT – II **(CO2) (15 Periods)**

Data Link Layer: Data Link Layer design issues, Error Detection and Correction, Elementary Data link Protocols, Sliding window protocols.

Medium Access Control Sublayer: The channel Allocation problem, Multiple Access Protocols, Ethernet.

UNIT – III **(CO3) (15 Periods)**

Network layer: Network layer Design Issues, Routing Algorithms - (The Optimality Principle, Shortest Path Routing, Flooding, Distance Vector Routing, Link State Routing, Hierarchical Routing, Broadcast Routing, Multicast Routing and Routing for Mobile Hosts). Congestion Control Algorithms, Quality of Service- Requirements, Techniques for Achieving Good Quality of Service. Internetworking, IPv4 Vs IPv6.

UNIT – IV **(CO4) (12 Periods)**

Transport Layer: Elements of Transport Protocols- – addressing: Connection Establishment, Connection Release, Error Control and Flow Control, UDP, RTP, TCP- Introduction to TCP, The TCP Service Model, The TCP Protocol, The TCP Segment Header, TCP Connection Establishment, TCP Connection Release.

Application Layer: DNS, Electronic Mail, HTTP, The World Wide Web - Architectural Overview.

Learning Resources:

Text Books:

1. Data Communication and Networking, Behrouz A. Forouzan, Tata McGraw-Hill, 4th Edition, 2010.
2. Computer Networks, A.S.Tanenbaum, 4th Edition, Pearson education.

Reference Books:

1. Introduction to Data communications and Networking, W.Tomasi, 4th Edition, Pearson Education.
2. Computer Networking A top down approach featuring the Internet, J.F.Kurose and K.W.Ross, Pearson Education, 5th Edition.
3. Computer Networks A Systems Approach, L.Peterson and B. Davie, Elsevier Morgan Kaufmann publisher, 5th Edition.

CD313	Machine Learning	L	T	P	C
		2	1	0	3

Course Objectives:

Explore the knowledge on

1. Basic concepts and applications of machine learning.
2. Supervised learning and its applications
3. Unsupervised learning and its applications
4. Multilayer perceptions and kernel tricks

Course Outcomes:

After successful completion of the course student shall be able to:

1. Apply the machine learning concepts in real life problems
2. Design solutions for supervised learning problems
3. Use rule sets and reinforcement learning to solve real world problems
4. Discuss the issues in dimensionality reduction and unsupervised learning algorithms.

Course Content:

UNIT – I **(CO1) (12 Periods)**

Introduction: Well posed learning problems, Designing a Learning System, Perspectives and Issues in machine learning.

Concept Learning and general to specific ordering: concept learning Task , Concept learning as a search, Finding a Maximally Specific Hypothesis , Version Spaces and Candidate Elimination Algorithm, Remarks on Version space and candidate elimination.

Bayesian Learning: Bayes Theorem, Maximum Likelihood and Least Square Error Hypotheses, Bayes Optimal Classifier, Naïve-Bayes Classifier, Bayesian Belief Network

UNIT – II **(CO2) (12 Periods)**

Decision Tree Learning : Decision Tree Representation, appropriate problems for decision tree, the basic decision tree Algorithm, Issues in decision tree learning.

Artificial Neural Networks: Introduction, Neural Network Representation, appropriate problems for neural network, Perceptrons , Multilayer Networks and the Back Propagation Algorithm.

Instance Based Learning: Introduction, KNN Learning, Locally Weighted Regression, Radial Bias Functions, Case-Based Reasoning.

UNIT – III **(CO3) (12 Periods)**

Learning Sets of Rules: Sequential Covering Algorithm, Learning Rule Sets: summary, Learning First Order Rules, Learning set of first order rules: FOIL.

Reinforcement Learning: Introduction, the Learning Task, Q Learning, Non Deterministic Rewards and Actions, Temporal Difference Learning , Generalizing from Examples , Relationship to Dynamic Programming.

UNIT – IV **(CO4) (12 Periods)**

Dimensionality Reduction: Introduction, subset selection, Principal component analysis, Feature Embedding, Factor analysis, Singular Value Decomposition and Matrix factorization, Multidimensional Scaling, Linear Discriminant analysis, Canonical correlation analysis.

Clustering: Introduction, Mixture Densities, K-Means Clustering, Expectation-Maximization Algorithm, Mixtures of Latent Variable Models, Supervised Learning after Clustering, Spectral Clustering, Hierarchical Clustering, Choosing the Number of Clusters.

Learning Resources:

Textbook:

1. Tom M. Mitchell, —Machine Learning, McGraw-Hill Education (India) Private Limited, 2013. (UNIT I, UNIT II, and UNIT III)
2. Ethem Alpaydin, Introduction to Machine Learning, MIT Press, Prentice Hall of India, Third Edition 2014. (UNIT IV)

Reference Books:

1. Stephen Marsland, —Machine learning: An Algorithmic Perspective‖, CRC Press, 2009.
2. Machine Learning: a Probabilistic Perspective, Kevin P. Murphy, MIT Press, 2012.
3. Foundations of Machine Learning, Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar, MIT Press, 2012.
4. Machine Learning -The Art and Science of Algorithms that Make Sense of Data, Peter Flach, Cambridge.

CD314	Distributed and Cloud Computing	L	T	P	C
		3	0	0	3

Course Objectives:

1. To learn about the concepts of distributed systems and distributed resource management.
2. To study the concepts cloud computing
3. To study about virtualization and cloud resource management
4. To understand the concepts of different platforms.

Course Outcomes:

On completion of the course, the students will be able to:

1. Appreciate distributed Computing, distributed resource management.
2. Develop the ability to understand and use the architecture of compute and storage cloud, service and delivery models.
3. Choose the appropriate technologies, algorithms and approaches for implementation and use of cloud.
4. Explain the core issues of cloud computing such as resource management and security.

Course Content:

UNIT – I **(CO1) (12 Periods)**

Introduction to Distributed System Concepts: Introduction to Distributed Systems-Characteristics-Issues in Distributed Systems-Distributed System Model-Request/Reply Protocols-RPC-RMI- Software Environments for Distributed Systems- Performance, Security and Energy Efficiency – Performance metrics, Fault tolerance, network threats – Energy efficiency in Distributed Computing

UNIT – II **(CO2) (13 Periods)**

Introduction to Cloud Computing:

Introduction to Cloud Computing, Evolution of Cloud Computing, Cloud Characteristics, Elasticity in cloud -On-demand Provisioning, NIST Cloud Computing

Reference Architecture, Architectural Design Challenges, Deployment Models: Public, Private and Hybrid Clouds, Service Models: Iaas, Paas, Saas, Benefits of Cloud Computing.

UNIT – III **(CO3) (14 Periods)**

Cloud Enabling Technologies:

Virtualization - Implementation levels of Virtualization – Levels, VMM, Design Requirements and Providers, Virtualization support at OS Level, Middleware Support for Virtualization – Full and Para Virtualization, CPU Virtualization, Memory Virtualization, I/O Virtualization.

Cloud Software and Computing Platforms: GAE(Google APP Engine),Programming Environment for GAE, Architecture of GFS, Amazon EC2,Amazon Web Services ,Microsoft Azure.

UNIT – IV **(CO4) (13 Periods)**

Cloud Management, Storage and Security:

Resource Provisioning Methods, Cloud Management Products, Cloud Storage, Provisioning Cloud Storage, Managed and Un Managed Cloud Storage

Cloud Security: Overview, Cloud Security Challenges, Security Architecture Design, Virtual Machine Security, Application Security, Data Security.

Learning Resources:

Text Books:

1. Andrew S.Tanennbaum,Maarten Van Steen,"Distributed Systems -Principles and Paradigms",Second Edition ,Pearson,2006
2. Kai Hwang Geoffrey C. Fox Jack J. Dongarra – "Distributed and Cloud Computing: From Parallel Processing to the Internet of Things".

CD315	Business Intelligence Tools	L	T	P	C
		3	0	0	3

Course Objectives:

1. To understand the concepts of Business intelligence.
2. To understand the relevance of Data Visualization in Business.
3. To provide hands on working with Tableau software.
4. To understand the methods of presentation, creating dashboards and animations.

Course Outcomes:

After successful completion of the course student shall be able to:

1. Apply visualization techniques for various data analysis tasks.
2. Presenting data through charts and maps.
3. Design information dashboard.

Course Content:

UNIT – I **(CO1) (12 Periods)**

Definition of BI – Historical Perspective of BI- Architecture of BI- Data Warehouse, Business Analytics, Business Performance Management, User Interface, Cyclical process of a business intelligence analysis.

Data and Information Visualization - A brief history of Data Visualization- Data Visualization for businesses- Different types of Charts- Business Activity Monitoring through Dashboard - Emergence of Data Visualization and Visual Analytics.

UNIT – II **(CO1) (12 Periods)**

Working with Tableau Data Source and Basic Charts: Introduction to Tableau, Connecting to Data Source: Text Files, Excel, Access, other databases, merging multiple data sources.

Univariate Charts – Creating tables, bar graphs, pie charts, histograms, line charts, stacked bar graphs, box plots, Showing aggregate measures.

Bivariate Charts – Creating tables, scatter plots, swapping rows and columns, adding trend lines, selecting color palettes, using dates.

Multivariate Charts and Maps: Facets, area charts, bullet graphs, dual axes charts, Gantt charts, heat maps, Maps – Setting geographical roles, placing marks on map, overlaying demographic data, choropleth maps, polygon shapes, customizing maps.

UNIT – III **(CO2) (12 Periods)**

User defined fields: Using predefined fields, calculating percentages, applying if-then logic, applying logical functions, showing totals and percentages, discretizing data, manipulating text, aggregate data.

UNIT – IV **(CO3) (12 Periods)**

Customization: Adding title and caption, font size and colors, adding various marks, adding reference lines, using presentation mode, adding annotation, adding drop-down selectors, search box selectors, slider selectors, creating dashboards, creating animated visualization.

Learning Resources:

Text Books:

1. Efraim Turban, Ramesh Sharda, Dursun Delen, "Decision Support and Business Intelligence Systems", Pearson.
2. Nandeshwar, A., Tableau Data Visualization Cookbook, Mumbai: PACKT / Shroff Publishers.
3. Iliinsky, N. & Steele, J., Designing Data Visualizations, Mumbai: O"Reilly / Shroff Publishers.
4. Milligan, N.J., Learning Tableau, Mumbai: PACKT / Shroff Publishers.
5. Jones, B., Communicating Data with Tableau, Mumbai: PACKT / Shroff Publishers.
6. Yau, N., Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics. John Wiley & Sons.