

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

CD 601- Deep Learning

COURSE OBJECTIVES: Introduce deep learning fundamentals and major algorithms, the problem settings, and their applications to solve real world problems.

COURSE OUTCOMES: After completing the course student should be able to:

1. Describe in-depth about theories, fundamentals, and techniques in Deep learning.
2. Identify the on-going research in computer vision and multimedia field.
3. Evaluate various deep networks using performance parameters.
4. Design and validate deep neural network as per requirements.

Unit I: Introduction History of Deep Learning, McCulloch Pitts Neuron, Multilayer Perceptions (MLPs), Representation Power of MLPs, Sigmoid Neurons, Feed Forward Neural Networks, Back propagation, weight initialization methods, Batch Normalization, Representation Learning, GPU implementation, Decomposition – PCA and SVD.

Unit II: Deep Feedforward Neural Networks, Gradient Descent (GD), Momentum Based GD, Nesterov Accelerated GD, Stochastic GD, AdaGrad, Adam, RMSProp, Auto-encoder, Regularization in auto-encoders, Denoising auto-encoders, Sparse auto-encoders, Contractive auto-encoders, Variational auto-encoder, Auto-encoders relationship with PCA and SVD, Dataset augmentation. Denoising auto encoders,

Unit III: Introduction to Convolutional neural Networks (CNN) and its architectures, CNN terminologies: ReLu activation function, Stride, padding, pooling, convolutions operations, Convolutional kernels, types of layers: Convolutional, pooling, fully connected, Visualizing CNN, CNN examples: LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet, RCNN etc. Deep Dream, Deep Art. Regularization: Dropout, drop Connect, unit pruning, stochastic pooling, artificial data, injecting noise in input, early stopping, Limit Number of parameters, Weight decay etc.

Unit IV: Introduction to Deep Recurrent Neural Networks and its architectures, Backpropagation Through Time (BPTT), Vanishing and Exploding Gradients, Truncated BPTT, Gated Recurrent Units (GRUs), Long Short Term Memory (LSTM), Solving the vanishing gradient problem with LSTMs, Encoding and decoding in RNN network, Attention Mechanism, Attention over images, Hierarchical Attention, Directed Graphical Models. Applications of Deep RNN in Image Processing, Natural Language Processing, Speech recognition, Video Analytics.

Unit V: Introduction to Deep Generative Models, Restricted Boltzmann Machines (RBMs), Gibbs Sampling for training RBMs, Deep belief networks, Markov Networks, Markov Chains, Auto-regressive Models: NADE, MADE, PixelRNN, Generative Adversarial Networks (GANs), Applications of Deep Learning in Object detection, speech/imagerecognition, video analysis, NLP, medical science etc.

TEXTBOOKS RECOMMENDED:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville; Deep Learning, MIT Press.
2. Charu C. Aggarwal "Neural Networks and Deep Learning: A Textbook", Springer.
3. Francois Chollet, "Deep Learning with Python", Manning Publications.

REFERENCE BOOKS:

1. Aurelien Geon, "Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems", O'Reilly.
2. Andreas Muller, "Introduction to Machine Learning with Python: A Guide for Data Scientists", O'Reilly.
3. Adam Gibson, Josh Patterson, "Deep Learning: A Practitioner's Approach", O'Reilly

LIST OF EXPERIMENTS:

1. Image Classification with CNN
2. Face Detection system with OpenCV library
3. Digit Recognition System with CNN
4. Music Genre Classification system (FMA: Free Music Archive Dataset)
5. Image Compression and De-compression using Encoders and Decoders
6. Predicting Airline Passengers count based on LSTM and RNN
7. Diabetes detection in patients with functional and sequential implementation of Keras
8. Detecting customer churn on banking dataset with Deep Neural Network
9. Multiclass wine classification using Neural Networks

Breast Cancer Detection using Neural Network Architecture

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

CD 602- Computer Networks

Course Outcomes: After completion of the course students will be able to

1. Characterize and appreciate computer networks from the viewpoint of components and from the viewpoint of services
2. Display good understanding of the flow of a protocol in general and a network protocol in particular
3. Model a problem or situation in terms of layering concept and map it to the TCP/IP stack
4. Select the most suitable Application Layer protocol (such as HTTP, FTP, SMTP, DNS, Bit torrent) as per the requirements of the network application and work with available tools to demonstrate the working of these protocols.
5. Design a Reliable Data Transfer Protocol and incrementally develop solutions for the requirements of Transport Layer
6. Describe the essential principles of Network Layers and use IP addressing to create subnets for any specific requirements

Unit –I

Computer Network: Definitions, goals, components, Architecture, Classifications & Types. Layered Architecture: Protocol hierarchy, Design Issues, Interfaces and Services, Connection Oriented & Connectionless Services, Service primitives, Design issues & its functionality. ISO/OSI Reference Model: Principle, Model, Descriptions of various layers and its comparison with TCP/IP. Principles of physical layer: Media, Bandwidth, Data rate and Modulations

Unit-II

Data Link Layer: Need, Services Provided, Framing, Flow Control, Error control. Data Link Layer Protocol: Elementary & Sliding Window protocol: 1-bit, Go-Back-N, Selective Repeat, Hybrid ARQ. Protocol verification: Finite State Machine Models & Petri net models. ARP/RARP/GARP

Unit-III

MAC Sub layer: MAC Addressing, Binary Exponential Back-off (BEB) Algorithm, Distributed Random Access Schemes/Contention Schemes: for Data Services (ALOHA and Slotted ALOHA), for Local-Area Networks (CSMA, CSMA/CD, CSMA/CA), Collision Free Protocols: Basic Bit Map, BRAP, Binary Count Down, MLMA Limited Contention Protocols: Adaptive Tree Walk, Performance Measuring Metrics. IEEE Standards 802 series & their variant.

Unit-IV

Network Layer: Need, Services Provided, Design issues, Routing algorithms: Least Cost Routing algorithm, Dijkstra's algorithm, Bellman-ford algorithm, Hierarchical Routing, Broadcast Routing, Multicast Routing. IP Addresses, Header format, Packet forwarding, Fragmentation, and reassembly, ICMP, Comparative study of IPv4 & IPv6

Unit-V

Transport Layer: Design Issues, UDP: Header Format, Per-Segment Checksum, Carrying Unicast/Multicast Real-Time Traffic, TCP: Connection Management, Reliability of Data Transfers, TCP Flow Control, TCP Congestion Control, TCP Header Format, TCP Timer Management. Application Layer: WWW and HTTP, FTP, SSH, Email (SMTP, MIME, IMAP), DNS, Network Management (SNMP).

References:

1. Andrew S. Tanenbaum, David J. Wetherall, "Computer Networks" Pearson Education.
2. Douglas E Comer, "Internetworking With Tcp/Ip Principles, Protocols, And Architecture - Volume I" 6th Edition, Pearson Education
3. Dimitri Bertsekas, Robert Gallager, "Data Networks", PHI Publication, Second Edition.
4. Kaveh Pahlavan, Prashant Krishnamurthy, "Networking Fundamentals", Wiley Publication.
5. Uyless Black, "Computer Networks", PHI Publication, Second Edition.
6. Ying-Dar Lin, Ren-Hung Hwang, Fred Baker, "Computer Networks: An Open Source Approach", McGraw Hill.

List of Experiments:

1. Study of Different Type of LAN & Network Equipments.
2. Study and Verification of standard Network topologies i.e. Star, Bus, Ring etc.
3. LAN installations and Configurations.
4. Write a program to implement various types of error correcting techniques.
5. Write a program to Implement various types of framing methods.
6. Study of Tool Command Language (TCL).
7. Study and Installation of Standard Network Simulator: N.S-2, N.S3, OpNet, QualNet etc .
8. Study & Installation of ONE (Opportunistic Network Environment) Simulator for High Mobility Networks .
9. Configure 802.11 WLAN.
10. Implement & Simulate various types of routing algorithm.
11. Study & Simulation of MAC Protocols like Aloha, CSMA, CSMA/CD and CSMA/CA using Standard Network Simulators.
12. Study of Application layer protocols-DNS, HTTP, HTTPS, FTP and TelNet.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

Departmental Elective CD603 (A) Big Data Analytics

Course Outcomes:

1. Students should be able to understand the concept and challenges of Big data.
2. Students should be able to demonstrate knowledge of big data analytics.
3. Students should be able to develop Big Data Solutions using Hadoop Eco System
4. Students should be able to gain hands-on experience on large-scale analytics tools.
5. Students should be able to analyse the social network graphs.

Course Content

Unit1: Introduction to Big data, Big data characteristics, Types of big data, Traditional versus Big data, Evolution of Big data, challenges with Big Data, Technologies available for Big Data, Infrastructure for Big data, Use of Data Analytics, Desired properties of Big Data system.

Unit2: Introduction to Hadoop, Core Hadoop components, Hadoop Eco system, Hadoop Physical Architecture, Hadoop limitations, RDBMS Versus Hadoop, Hadoop Distributed Filesystem, Processing Data with Hadoop, Managing Resources and Application with Hadoop YARN, MapReduce programming.

Unit3: Introduction to Hive Hive Architecture, Hive Data types, Hive Query Language, Introduction to Pig, Anatomy of Pig, Pig on Hadoop, Use Case for Pig, ETL Processing, Datatypes in Pig running Pig, Execution model of Pig, Operators, functions, Data types of Pig.

Unit4: Introduction to NoSQL, NoSQL Business Drivers, NoSQL Data architectural patterns, Variations of NoSQL architectural patterns using NoSQL to Manage Big Data, Introduction to MongoDB

Unit5: Mining social Network Graphs: Introduction Applications of social Network mining, Social Networks as a Graph, Types of social Networks, Clustering of social Graphs, Direct Discovery of communities in a social graph, Introduction to recommender system.

Text Books:

1. Radha Shankarmani, M. Vijayalakshmi, " Big Data Analytics", Wiley, Second edition
2. Seema Acharya, Subhashini Chellappan, " Big Data and Analytics", Wiley, First edition

Reference Books:

1. 1.KaiHwang,Geoffrey C., Fox. Jack, J. Dongarra, “Distributed and Cloud Computing”, Elsevier, Firstedition
2. Michael Minelli, Michele Chambers, AmbigaDhiraj, “Big Data Big Analytics”, Wileyfor old question papers visit <http://www.rgpvonline.com>

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

Departmental Elective CD603 (B) Data Acquisition

Unit I - Introduction to Digital Data Systems

Digital Data, Data Acquisition, Methods of Digital Data Systems, Control of Digital Data System, Cyber Crime Investigations, Analysis of Cyber Crime, Cyber Ethics and Landscapes.

Unit II - Data Analysis

Cybercheck Analyse True Back Image, Encase Image, Raw Disk Dumps, Virtual disk images and RAM, Dumps, Investigations on Report Analysis, cyber Laws in Secure Analysis, Extract unallocated and disk slack areas, Data carving of slack areas, Hash Files, File Signature.

Unit III - File System

File Systems CyberCheck FAT12/16/32, exFAT, NTFS, Linux EXT2/3/4 FS, UFS, CDFS, Sun Solaris, Reiser FS, Unix(Free BSD) and MAC, Dynamic Disks and Linux RAID Disks. Analysis - Introduction, File system category, Content category, Metadata category, File name category, The big picture, File recovery, determining the type, Consistency check, FAT data structure directory entries.

Unit IV - Data Recovery

Cybercheck Recover Data, Deleted files/folders Recovery, Deleted Partitions, Data Formatted Partition Data, Add Data, Report Data Findings.

Unit V - Digital Communication

Introduction to Digital Communication, Input/ Output Interfacing, Digital Inputs, Data Isolation, Data Sheets, Smart Networks, Network Communication System Architecture, Wireless Networks, Network Protocols, Data Transmission.

TEXT BOOKS:

1. Parasram, S. V. N. (2020), "Digital Forensics with Kali Linux" - Second Edition: Perform Data Acquisition, Data Recovery, Network Forensics, and Malware Analysis with Kali Linux. United Kingdom: Packt Publishing, Limited.
2. Carrier, B. (2005). "File System Forensic Analysis", United Kingdom: Pearson Education.

REFERENCE BOOKS:

1. I.Kruse, W. G., Heiser, J. G. (2001). Computer Forensics: Incident Response Essentials. (n.p.): Pearson Education.
2. Parasram, S. V. N., Joseph, D., Samm, A. (2017). Digital Forensics with Kali Linux: Perform Data Acquisition, Digital Investigation, and Threat Analysis Using Kali Linux Tools. United Kingdom: Packt Publishing.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

Departmental Elective CD603 (C) Advanced Database Management System

UNIT-I

Objected Oriented and Object Relational Databases Modeling Complex Data Semantics, Specialization, Generalization, Aggregation and Association, Objects, Object Identity and its implementation, Clustering, Equality and Object Reference, Architecture of Object Oriented and Object Relational databases, Persistent Programming Languages, Cache Coherence. Case Studies: Gemstone, O2, Object Store, SQL3, Oracle xxi, DB2.

UNIT-II

Deductive Databases Data log and Recursion, Evaluation of Data log program, Recursive queries with negation. Parallel and Distributed Databases Parallel architectures, shared nothing/shared disk/shared memory based architectures, Data partitioning, Intra-operator parallelism, pipelining. Distributed Data Storage – Fragmentation & Replication, Location and Fragment Transparency Distributed Query Processing and Optimization, Distributed Transaction Modeling and concurrency Control, Distributed Deadlock, Commit Protocols, Design of Parallel Databases, and Parallel Query Evaluation.

UNIT-III

Advanced Transaction Processing Advanced transaction models: Savepoints, Nested and Multilevel Transactions, Compensating Transactions and Saga, Long Duration Transactions, Weak Levels of Consistency, Transaction Work Flows, Transaction Processing Monitors, Shared disk systems.

UNIT-IV

Active Database and Real Time Databases Triggers in SQL, Event Constraint and Action: ECA Rules, Query Processing and Concurrency control, Recursive query processing, Compensation and Databases Recovery, multi-level recovery.

UNIT-V

Image and Multimedia Databases Modeling and Storage of Image and Multimedia Data, Data Structures – R-tree, k-d tree, Quad trees, Content Based Retrieval: Color Histograms, Textures, etc., Image Features, Spatial and Topological Relationships, Multimedia Data Formats, Video Data Model, Audio & Handwritten Data, Geographic Information Systems (GIS).

WEB Database

Accessing Databases through WEB, WEB Servers, XML Databases, Commercial Systems – Oracle xxi, DB2.

BOOKS

1. Elmarsi, “Fundamentals of Database Systems”, 4 th Edition, Pearson Education
2. R. Ramakrishnan, “Database Management Systems”, 1998, McGraw Hill International Editions
3. Elmagarmid.A.K. “Database transaction models for advanced applications”, Morgan Kaufman.
4. Transaction Processing, Concepts and Techniques, J. Gray and A. Reuter, Morgan Kauffman..
5. S. Abiteboul, R. hull and V. Vianu, “Foundations of Databases”, 1995, Addison – Wesley Publishing Co., Reading Massachusetts.
6. W. Kim, “Modern Database Systems”, 1995, ACM Press, Addison – Wesley.
D. Maier, “The Theory of Relational Databases”, 1993, Computer Science Press, Rockville, Maryland

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

Open Elective CD604 (A) Information Extraction and Retrieval

UNIT-I

Introduction - History of IR- Components of IR - Issues -Open source Search engine Frameworks - The Impact of the web on IR - The role of artificial intelligence (AI) in IR – IR Versus Web Search - Components of a search engine, Characterizing the web.

UNIT -II

Boolean and Vector space retrieval models- Term weighting - TF-IDF weighting- cosine similarity - Preprocessing - Inverted indices - efficient processing with sparse vectors Language Model based IR - Probabilistic IR -Latent Semantic indexing - Relevance feedback and query expansion.

UNIT- III

Web search overview, web structure the user paid placement search engine optimization, Web Search Architectures - crawling - meta-crawlers, Focused Crawling - web indexes - Nearduplicate detection - Index Compression - XML retrieval.

UNIT -IV

Link Analysis -hubs and authorities - Page Rank and HITS algorithms -Searching and Ranking - Relevance Scoring and ranking for Web - Similarity - Hadoop & Map Reduce - Evaluation - Personalized search - Collaborative filtering and content-based recommendation of documents And products - handling invisible Web - Snippet generation, Summarization. Question Answering, Cross- Lingual Retrieval.

UNIT -V

Information filtering: organization and relevance feedback - Text Mining- Text classification and clustering - Categorization algorithms, naive Bayes, decision trees and nearest neighbor - Clustering algorithms: agglomerative clustering, k-means, expectation maximization (EM).

References:

1. C. Manning, P. Raghvan and H Schutze: Introduction to Information Retrieval, Cambridge University Press, 2008.
2. Ricardo Baeza -Yates and Berthier Ribeiro –Neto, Modern Information Retrieval. The Concepts and Technology behind Search 2nd Edition, ACM Press Books 2011.
3. Bruce Croft, Donald Metzler and Trevor Strohman Search Engines Information Retrieval in Practice 1st Edition Addison Wesley, 2009
4. 4.Mark Levene, An Introduction to Search Engines and Web Navigation, 2nd Edition Wiley 2010.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

Open Elective CD604 (B) Agile Software Development

Pre-Requisite: Software Engineering

Course Outcomes:

After completing the course student should be able to:

1. Describe the fundamental principles and practices associated with each of the agile development methods.
2. Compare agile software development model with traditional development models and identify the benefits and pitfalls.
3. Use techniques and skills to establish and mentor Agile Teams for effective software development.
4. Apply core values and principles of Agile Methods in software development.

Course Contents:

Unit-I: Fundamentals of Agile Process: Introduction and background, Agile Manifesto and Principles, Stakeholders and Challenges, Overview of Agile Development Models: Scrum, Extreme Programming, Feature Driven Development, Crystal, Kanban, and Lean Software Development.

Unit-II: Agile Projects: Planning for Agile Teams: Scrum Teams, XP Teams, General Agile Teams, Team Distribution; Agile Project Lifecycles: Typical Agile Project Lifecycles, Phase Activities, Product Vision, Release Planning: Creating the Product Backlog, User Stories, Prioritizing and Estimating, Creating the Release Plan; Monitoring and Adapting: Managing Risks and Issues, Retrospectives.

Unit-III: Introduction to Scrum: Agile Scrum Framework, Scrum Artifacts, Meetings, Activities and Roles, Scrum Team Simulation, Scrum Planning Principles, Product and Release Planning, Sprinting: Planning, Execution, Review and Retrospective; User story definition and Characteristics, Acceptance tests and Verifying stories, Burn down chart, Daily scrum, Scrum Case Study.

Unit-IV: Introduction to Extreme Programming (XP): XP Lifecycle, The XP Team, XP Concepts: Refactoring, Technical Debt, Timeboxing, Stories, Velocity; Adopting XP: Pre-requisites, Challenges; Applying XP: Thinking- Pair Programming, Collaborating, Release, Planning, Development; XP Case Study.

Unit-V: Agile Software Design and Development: Agile design practices, Role of design Principles, Need and significance of Refactoring, Refactoring Techniques, Continuous Integration, Automated build tools, Version control; Agility and Quality Assurance: Agile Interaction Design, Agile approach to Quality Assurance, Test Driven Development, Pair

programming: Issues and Challenges.

Recommended Books:

1. Robert C. Martin, Agile Software Development- Principles, Patterns and Practices, Prentice Hall, 2013.
2. Kenneth S. Rubin, Essential Scrum: A Practical Guide to the Most Popular Agile Process, Addison Wesley, 2012.
3. James Shore and Shane Warden, The Art of Agile Development, O'Reilly Media, 2007.
4. Craig Larman, —Agile and Iterative Development: A manager's Guide, Addison-Wesley, 2004.
5. Ken Schawber, Mike Beedle, Agile Software Development with Scrum, Pearson, 2001.
6. Cohn, Mike, Agile Estimating and Planning, Pearson Education, 2006.
7. Cohn, Mike, User Stories Applied: For Agile Software Development Addison Wisley, 2004.

Online Resources:

1. IEEE Transactions on Software Engineering
2. IEEE Transactions on Dependable and Secure Computing
3. IET Software
4. ACM Transactions on Software Engineering and Methodology (TOSEM)
5. ACM SIGSOFT Software Engineering Notes

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE-Data Science/Data Science, VI semester

Open Elective CD604 (C) Natural Language Processing

COURSE OBJECTIVES: Students should develop a basic understanding in natural language processing methods and strategies and to evaluate the strengths and weaknesses of various Natural Language Processing (NLP) methods & technologies and gain an insight into the application areas of Natural language processing.

Detailed Contents:

UNIT I: Introduction: Origins and challenges of NLP, Human languages, Application and Goals of NLP, Main approach of NLP, Knowledge in speech and language processing, Ambiguity, Models and algorithms, Formal language and Natural Language, Regular Expression, and automata.

UNIT II: Text Pre-processing, Tokenization, Feature Extraction from text Morphology: Inflectional morphology, Derivational morphology, Finite state morphological parsing, Morphology, and Indian languages. Part of Speech Tagging: Rule based, Stochastic POS, HMM, Transformation based tagging (TBL), N-Grams: Simple N-grams, Smoothing, Backoff, Entropy. Handling of unknown words, Named entities, Multi word expressions.

UNIT III: Parsing: Syntactic and statistical parsing, parsing algorithms, hybrid of rule based and probabilistic parsing, scope ambiguity and attachment ambiguity resolution, Tree banks. Discourse and dialogue: discourse and dialogue analysis, anaphora resolution, named entity resolution, event anaphora, Information extraction and retrieval. Hidden Markov and Maximum Entropy models, Viterbi algorithms and EM training.

UNIT IV: Semantic analysis, Semantic attachments – Word Senses, Relations between Senses, Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods. Compositional semantics.

Speech Processing: Speech and phonetics, Vocal organ, Phonological rules and Transducer, Probabilistic models: Spelling error, Bayesian method to spelling, Minimum edit distance, Bayesian method of pronunciation variation.

UNIT V: Application of NLP: intelligent work processors: Machine translation, user interfaces, Man-Machine interfaces, natural language querying, tutoring and authoring systems, speech recognition, and commercial use of NLP.

Text Books:

1. Daniel Jurafsky, James H. Martin—Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, Pearson Publication.
2. Steven Bird, Ewan Klein and Edward Loper, —Natural Language Processing with Python, OReilly Media.
3. Manning and Schutze "Foundations of Statistical Natural Language Processing", MIT Press.

Reference Books:

1. Breck Baldwin, Language Processing with Java and LingPipe Cookbook, Atlantic Publisher.
2. Richard M Reese, Natural Language Processing with Java, OReilly Media.
3. Nitin Indurkha and Fred J. Damerau, Handbook of Natural Language Processing, Chapman and Hall/CRC Press.
4. Tanveer Siddiqui, U.S. Tiwary, Natural Language Processing and Information Retrieval, Oxford University Press.