RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD701 Data Engineering

UNIT – I

Data Driven Organizations & Elements of Data:

Data-driven decisions, data pipeline infrastructure for data-driven decisions, role of the data engineer in data-driven organizations, Modern data strategies, Introduction to elements of Data, the five Vs of data – volume, velocity, variety, veracity, and value, Variety – data types & data sources, Activities to improve veracity and value.

UNIT – II

Design Principles and Patterns for Data Pipelines

The evolution of data architectures, Modern data architecture on various cloud platforms, Modern data architecture pipeline - Ingestion and storage, Modern data architecture pipeline - Processing and Consumption, Streaming analytics pipeline

Securing and Scaling the Data Pipeline:

Cloud security, Security of analytics workloads, ML security, Scaling Data Pipeline, creating a scalable infrastructure, Creating scalable components.

UNIT – III

Ingesting and Preparing Data:

ETL and ELT comparison, Data wrangling, Data Discovery, Data structuring, Data Cleaning, Data enriching, Data validating, Data publishing

Ingesting by Batch or by Stream

Comparing batch and stream ingestion, Batch ingestion processing, Purpose-built data ingestion tools, Scaling considerations for batch processing, stream processing, Scaling considerations for stream processing, Ingesting IoT data by stream

UNIT – IV

Storing and Organizing Data

Storage in the modern data architecture, Data Lake storage, Data warehouse storage, Purpose-built databases, Storage in support of the pipeline, Securing storage.

Processing Big Data

Big data processing concepts, Apache Hadoop, Apache Spark, Amazon EMR

UNIT – V

Processing Data for ML & Automating the Pipeline:

ML Concepts, ML Lifecycle, Framing the ML problem to meet the business goal, Collecting data, Applying labels to training data with known targets, Pre-processing data, Feature engineering, Developing a model, Deploying a model, ML infrastructure on AWS, AWS SageMaker, Automating the Pipeline, Automating infrastructure deployment, CI/CD, Automating with Step Functions.

List of Experiments:

- 7 - 10 experiments to be framed as per the syllabus.

Recommended Books:

1. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems, by Martin Kleppmann
2. T-SQL Querying (Developer Reference) by Itzik Ben-Gan
3. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling by Margy Ross
4. Spark: The Definitive Guide: Big Data Processing Made Simple by Bill Chambers
5. Data Pipelines with Apache Airflow by Bas P. Harenslak
6. Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing by Tyler Akidau
7. Kubernetes in Action by Marko Luksa

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD702 (A) Data Analytics & Visualization

Unit 1: Data Definitions and Analysis Techniques: Elements, Variables, and Data categorization Levels of Measurement Data management and indexing Introduction to Statistical Concepts: Sampling Distributions, Resampling, Statistical Inference and Descriptive Statistics, Measures of central tendency, Measures of location of dispersions

Unit 2: Advance Data analysis techniques: Statistical hypothesis generation and testing, Chi-Square test, t-Test, Analysis of variance, Correlation analysis, Maximum likelihood test, Regression Modelling, Multivariate Analysis, Bayesian Modelling, Inference and Bayesian Network, Regression analysis

Unit 3: Data Wrangling: Intro to Data Wrangling, Gathering Data, Assessing Data, Cleaning Data. Data Visualization in Data Analysis: Design of Visualizations, Univariate Exploration of Data, Bivariate Exploration of Data, Multivariate Exploration of Data, Explanatory Visualizations.

Unit 4: Data Ecosystem: Overview of the Data Analyst Ecosystem, Types of Data, Understanding Different Types of File Formats, Sources of Data, Overview of Data Repositories, NoSQL, Data Marts, Data Lakes, ETL, and Data Pipelines, Foundations of Big Data, Big Data processing tools such as Hadoop, Hadoop Distributed File System (HDFS), Hive, and Spark

Unit 5: Data Visualization tools: Python visualization libraries (matplotlib, pandas, seaborn, ggplot, plotly), Introduction to PowerBI tools, Examples of inspiring (industry) projects- Exercise: create your own visualization of a complex dataset.

Text Books/References:

1. Joel Grus, Data Science from Scratch, Shroff Publisher Publisher /O'Reilly Publisher Media
2. Annalyn Ng, Kenneth Soo, Numsense! Data Science for the Layman, Shroff Publisher Publisher
3. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly Publisher Media.
4. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press.
5. Jake VanderPlas, Python Data Science Handbook, Shroff Publisher Publisher /O'Reilly Publisher Media 6. Philipp Janert, Data Analysis with Open Source Tools, Shroff Publisher Publisher /O'Reilly Publisher Media.

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD702 (B) Wireless and Mobile Computing

Unit-1

Review of traditional networks: Review of LAN, MAN, WAN, Intranet, Internet, and interconnectivity devices: bridges, Routers, etc. Review of TCP/IP Protocol Architecture: ARP/RARP, IP addressing, IP Datagram format and its Delivery, Routing table format, ICMP Messages, Subnetting, Super netting, and CIDR, DNS. NAT: Private addressing and NAT, SNAT, DNAT, NAT and firewalls, VLANS: Concepts, Comparison with Real LANS, Type of VLAN, Tagging, IPV6: address structure, address space and header.

Unit 2

Study of traditional routing and transport: Routing Protocols: BGP- Concept of hidden network and autonomous system, An Exterior gateway protocol, Different messages of BGP. Interior Gateway protocol: RIP, OSPF. Multiplexing and ports, TCP: Segment format, Sockets, Synchronization, Three Way Hand Shaking, Variable window size and Flow control, Timeout and Retransmission algorithms, Connection Control, Silly window Syndrome. Example of TCP: Taho, Reno, Sack etc. UDP: Message Encapsulation, Format and Pseudo header.

Unit 3

Wireless LAN: Transmission Medium For WLANs, MAC problems, Hidden and Exposed terminals, Near and Far terminals, Infrastructure and Ad hoc Networks, IEEE 802.11- System arch, Protocol arch, Physical layer, Concept of spread spectrum, MAC and its management, Power management, Security. Mobile IP: unsuitability of Traditional IP; Goals, Terminology, Agent advertisement and discovery, Registration, Tunneling techniques. Ad hoc network routing: Ad hoc Network routing v/s Traditional IP routing, types of routing protocols, Examples: OADV, DSDV, DSR, ZRP etc.

Unit 4

Mobile transport layer: unsuitability of Traditional TCP; I-TCP, S-TCP, M-TCP. Wireless Cellular networks: Cellular system, Cellular networks v/s WLAN, GSM – Services, system architecture, Localization and calling, handover and Roaming.

Unit 5

Mobile Device Operating Systems: Special Constraints & Requirements, Commercial Mobile Operating Systems. Software Development Kit: iOS, Android etc. MCommerce: Structure, Pros &Cons, Mobile Payment System, Security Issues

TEXTBOOKS RECOMMENDED:

1. Comer, “Internetworking with TCP/ IP Vol-I”, 5th edition, Addison Wesley, 2006.
2. Jochen Schiller “Mobile communication”, 2nd edition, Pearson education, 2008

REFERENCE:

1. W. Richard Stevens, “TCP/IP Illustrated Vol-I”, Addison-Wesley.
2. C.K.Toh, “AdHoc Mobile Wireless Networks”, First Edition, Pearson Education.
3. Uwe Hansmann, LotharMerk, Martin S. Nicklons and Thomas Stober, “Principles of Mobile Computing”, Springer

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD702 (C) Cloud Computing

Course Objective: The objective of this course is to provide students with comprehensive and in depth knowledge of Cloud Computing concepts, technologies, architecture and applications.

UNIT I Introduction of Grid and Cloud computing, characteristics, components, business and IT perspective, cloud services requirements, cloud models, Security in public model, public verses private clouds, Cloud computing platforms: Amazon EC2,Platform as Service: Google App Engine, Microsoft Azure, Utility Computing, Elastic Computing.

UNIT II Cloud services- SAAS, PAAS, IAAS, cloud design and implementation using SOA, conceptual cloud model, cloud stack, computing on demand, Information life cycle management, cloud analytics, information security, virtual desktop infrastructure, storage cloud.

UNIT III Virtualization technology: Definition, benefits, sensor virtualization, HVM, study of hypervisor, logical partitioning- LPAR, Storage virtualization, SAN, NAS, cloud server virtualization, virtualized data center.

UNIT IV Cloud security fundamentals, Vulnerability assessment tool for cloud, Privacy and Security in cloud, Cloud computing security architecture: Architectural Considerations- General Issues, Trusted Cloud computing, Secure Execution Environments and Communications, Microarchitectures; Identity Management and Access control-Identity management, Access control, Autonomic Security, Cloud computing security challenges: Virtualization security managementvirtual threats, VM Security Recommendations, VM-Specific Security techniques, Secure Execution Environments and Communications in cloud.

UNIT V SOA and cloud, SOA and IAAS, cloud infrastructure benchmarks, OLAP, business intelligence, e-Business, ISV, Cloud performance monitoring commands, issues in cloud computing. QOS issues in cloud, mobile cloud computing, Inter cloud issues, Sky computing, Cloud Computing Platform, Xen Cloud Platform, Eucalyptus, OpenNebula, Nimbus, TPlatform, Apache Virtual Computing Lab (VCL), Anomaly Elastic Computing Platform.

References:

1. Dr.Kumar Saurabh, “Cloud Computing”, Wiley India.
2. Ronald Krutz and Russell Dean Vines, “Cloud Security”, Wiley-India.
3. Judith Hurwitz, R.Bloor, M.Kanfman, F.Halper, “Computing for Dummies”, Wiley India Edition.
4. Anthony T.Velte Toby J.Velte, “Cloud Computing – A Practical Approach”, TMH.
5. Barrie Sosinsky, ‘Cloud Computing Bible”, Wiley India.

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

1. Explain the core concepts of the cloud computing paradigm
2. Demonstrate knowledge of virtualization
3. Explain the core issues of cloud computing such as security, privacy, and interoperability.
4. Choose the appropriate technologies, algorithms, and approaches for the related issues.
5. Identify problems, and explain, analyze, and evaluate various cloud computing solutions

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD704 (A) Advanced statistics for Data Science

Unit 1: Test of hypothesis: Concept & formulation, Type I and Type II errors, Neyman Pearson lemma, Procedures of testing, Basics of Time Series Analysis & Forecasting: Stationary, ARIMA Models: Identification, Estimation and Forecasting.

Unit 2: Non-parametric Inference: Comparison with parametric inference, Use of order statistics. Sign test, Wilcoxon signed rank test, Mann-Whitney test, Run test, Kolmogorov-Smirnov test. Spearman‘s and Kendall‘s test. Tolerance region.

Unit 3:Greatest Common Divisors and Prime Factorization: Greatest common divisors, The Euclidean algorithm, The fundamental theorem of arithmetic, Factorization of integers and the Fermat numbers Congruence: Introduction to congruence’s, Linear congruences, The Chinese remainder theorem, Systems of linear congruences

Unit 4: Stochastic Processes and Markov Chains: Introduction to Stochastic processesMarkov process. Transition Probability, Transition Probability Matrix, First order and Higher order Markov process, n step transition probabilities, Markov chain, Steady state condition, Markov analysis.

Unit 5: R statistical programming language: Introduction to R, Functions, Control flow and Loops, Working with Vectors and Matrices, Reading in Data, Writing Data, Working with Data, Manipulating Data, Simulation, Linear model, Data Frame, Graphics in R

TEXT BOOKS:

1 .S. D. Sharma, Operations Research, Kedarnath and Ramnath Publishers, Meerut, Delhi
2. . S C Gupta and V K Kapoor, Fundamentals of Mathematical statistics, Khanna publications
3. A. Goon, M. Gupta and B. Dasgupta, ―Fundamentals of Statistics, vol. I & II, World Press
4. Garrett Grolemund, “Hands-on Programming with R”, 1st edition, O′Reilly. Jared P. Lander, “R for Everyone: Advanced Analytics and Graphics”, 2nd edition, Addison-Wesley Professional.
5. SP GUPTA,Statistical Methods 48TH EDITION

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD703 (A) Cryptography & Information Security

Unit 1

Security Concepts: Introduction, The need for security, Security approaches, Principles of security, Types of Security attacks, Security services, Security Mechanisms, A model for Network Security, Cryptography Concepts and Techniques: Introduction, plain text and cipher text, substitution techniques, transposition techniques, encryption and decryption, symmetric and asymmetric key cryptography, steganography, key range and key size, possible types of attacks.

Unit 2

Symmetric key Ciphers: Block Cipher principles, DES, AES, Blowfish, RC5, IDEA, Block cipher operation, Stream ciphers, RC4. Asymmetric key Ciphers: Principles of public key cryptosystems, RSA algorithm, Elgamal Cryptography, Diffie-Hellman Key Exchange, Knapsack Algorithm.

Unit 3

Cryptographic Hash Functions: Message Authentication, Secure Hash Algorithm (SHA-512), Message authentication codes: Authentication requirements, HMAC, CMAC, Digital signatures, Elgamal Digital Signature Scheme. Key Management and Distribution: Symmetric Key Distribution Using Symmetric & Asymmetric Encryption, Distribution of Public Keys, Kerberos, X.509 Authentication Service, Public – Key Infrastructure.

Unit 4

Transport-level Security: Web security considerations, Secure Socket Layer and Transport Layer Security, HTTPS, Secure Shell (SSH) Wireless Network Security: Wireless Security, Mobile Device Security, IEEE 802.11 Wireless LAN, IEEE 802.11i Wireless LAN Security.

Unit 5

E-Mail Security: Pretty Good Privacy, S/MIME IP Security: IP Security overview, IP Security architecture, Authentication Header, Encapsulating security payload, Combining security associations, Internet Key Exchange Case Studies on Cryptography and security: Secure Multiparty Calculation, Virtual Elections, Single sign On, Secure Inter-branch Payment Transactions, Cross site Scripting Vulnerability.

TEXT BOOKS:

1. Cryptography and Network Security - Principles and Practice: William Stallings, Pearson Education, 6th Edition
2. Cryptography and Network Security: Atul Kahate, Mc Graw Hill, 3rd Edition

REFERENCE BOOKS:

1. Cryptography and Network Security: C K Shyamala, N Harini, Dr T R Padmanabhan, Wiley India, 1st Edition.
2. Cryptography and Network Security: Forouzan Mukhopadhyay, Mc Graw Hill, 3rd Edition
3. Information Security, Principles, and Practice: Mark Stamp, Wiley India.
4. Principles of Computer Security: WM. Arthur Conklin, Greg White, TMH
5. Introduction to Network Security: Neal Krawetz, CENGAGE Learning
6. Network Security and Cryptography: Bernard Menezes, CENGAGE Learning

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD703 (B) Intelligent Systems for Robotics

Course Objective:

The students will be able to understand the basic concepts and fundamentals of robotics. They will also be able to use AI in the field of robotics.

Detailed Contents:

Unit 1:

Introduction: Introduction to Robotics Fundamentals of Robotics, Robot Kinematics: Position Analysis, Dynamic Analysis and Forces, Robot Programming languages & systems: Introduction, the three levels of robot programming, requirements of a robot programming language, problems peculiar to robot programming languages.

Unit 2:

Need of AI in Robotics: History, state of the art, Need for AI in Robotics. Thinking and acting humanly, intelligent agents, structure of agents.

Unit 3:

Game Playing: AI and game playing, plausible move generator, static evaluation move generator, game playing strategies, problems in game playing.

Unit 4:

Robotics fundamentals: Robot Classification, Robot Specification, notation, kinematic representations and transformations, dynamics techniques; trajectory planning and control.

Unit 5:

Robotics and Its applications: DDD concept, Intelligent robots, Robot anatomy-Definition, law of robotics, History and Terminology of Robotics-Accuracy and repeatability of Robotics-Simple problems-Specifications of Robot-Speed of Robot, Robot joints and links-Robot classifications Architecture of robotic systems-Robot Drive systems-Hydraulic, Pneumatic and Electric system

Suggested References:

1. Robotics, Vision and Control: Fundamental Algorithms in MATLAB, Peter Corke, Springer, 2011.
2. Robotics: Everything You Need to Know About Robotics from Beginner to Expert, Peter McKinnon, Createspace Independent Publishing Platform, 2016.
3. Introduction to AI Robotics, Second Edition, By Robin R. Murphy, MIT press, 2001.
4. Artificial Intelligence for Robotics: Build intelligent robots that perform human tasks using AI techniques, Francis X. Govers, Packt Publishers, 2018

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD703 (C) Computer Vision

Unit 1
Introduction to Computer Vision and Basic Concepts of Image Formation: Introduction and Goals of Computer Vision and Image Processing, Image Formation Concepts. Fundamental Concepts of Image Formation: Radiometry, Geometric Transformations, Geometric Camera Models.

Unit-2
Fundamental Concepts of Image Formation: Camera Calibration, Image Formation in a Stereo Vision Setup, Image Reconstruction from a Series of Projections. Image Processing Concepts: Image Transforms.

Unit 3
Image Processing Concepts: Image Transforms, Image Enhancement. Image Processing Concepts: Image Filtering, Colour Image Processing, Image Segmentation. Image Descriptors and Features: Texture Descriptors, Colour Features, Edges/Boundaries.

Unit 4.
Image Descriptors and Features: Object Boundary and Shape Representations. Image Descriptors and Features: Interest or Corner Point Detectors, Histogram of Oriented Gradients, Scale Invariant Feature Transform, Speeded up Robust Features, Saliency

Unit 5
Applications of Computer Vision: Artificial Neural Network for Pattern Classification, Convolutional Neural Networks, Autoencoders. Applications of Computer Vision: Gesture Recognition, Motion Estimation and Object Tracking, Programming Assignments.

Books and references

1. Forsyth & Ponce, “Computer Vision-A Modern Approach”, Pearson Education.
2. M.K. Bhuyan , “ Computer Vision and Image Processing: Fundamentals and Applications”, CRC Press, USA, ISBN 9780815370840 - CAT# K338147.
3. Richard Szeliski, “Computer Vision- Algorithms & Applications”, Springer.

Online Lectures links
https://onlinecourses.nptel.ac.in/noc23_ee39/preview
https://onlinecourses.nptel.ac.in/noc19_cs58/preview
https://onlinecourses.nptel.ac.in/noc23_ee78/preview

Lab experiments
4. Write a python program for image enhancement.
5. Write a python program to perform compression operation on the input image.
6. Write a python program for color image processing on the input image.
7. Write a python program to perform image segmentation operation.
8. Write a python program to perform image morphology operation on the image.
9. Write a python program for Image Restoration operation.
10. Write a python program to implement Scaling, Rotating, Shifting and Edge Detection operations on input image.
11. Write a program for object tracking using Open CV

RAJIV GANDHI PROUDYOGIKI VISHWAVIDYALAYA, BHOPAL

New Scheme Based On AICTE Flexible Curricula

CSE- Data Science/Data Science, VII-Semester

CD703 (D) Data Mining and Warehousing

COURSE OBJECTIVES

Students should understand the value of Historical data and data mining in solving real-world problems.

Students should become affluent with the basic Supervised and unsupervised learning algorithms commonly used in data mining .

Students develop the skill in using data mining for solving real-world problems.

Unit 1. Data Warehousing: Introduction, Delivery Process, Data warehouse Architecture, Data Preprocessing: Data cleaning, Data Integration and transformation, Data reduction. Data warehouse Design: Data Warehouse schema, Partitioning strategy Data warehouse Implementation, Data Marts, Meta Data, Example of a Multidimensional Data model. Introduction to Pattern Warehousing.

Unit 2. OLAP Systems: Basic concepts, OLAP queries, Types of OLAP servers, OLAP operations etc. Data Warehouse Hardware and Operational Design: Security, Backup And Recovery,

Unit 3. Introduction to Data & Data Mining :Data Types, Quality of data, Data Preprocessing, Similarity measures, Summary statistics, Data distributions, Basic data mining tasks, Data Mining V/s knowledge discovery in databases. Issues in Data mining. Introduction to Fuzzy sets and fuzzy logic.

Unit 4. Supervised Learning: Classification: Statistical-based algorithms, Distance-based algorithms, Decision tree-based algorithms, Neural network-based algorithms, Rule-based algorithms, Probabilistic Classifiers

Unit 5. Clustering & Association Rule mining : Hierarchical algorithms, Partitional algorithms, Clustering large databases – BIRCH, DBSCAN, CURE algorithms.Association rules : Parallel and distributed algorithms such as Apriori and FP growth algorithms.

Books Recommended: Text Books:

1. Pang – ningTan , Steinbach & Kumar, “Introduction to Data Mining”, Pearson Edu, 2019.
2. Jaiwei Han, Micheline Kamber, “Data Mining : Concepts and Techniques”, Morgan Kaufmann Publishers.

Reference Books:

1. Margaret H. Dunham, “Data Mining : Introductory and Advanced topics”, Pearson Edu., 2009.
2. Anahory& Murray, “Data Warehousing in the Real World”, Pearson Edu., 2009.