JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B.Tech. in COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)
COURSE STRUCTURE & SYLLABUS (R22 Regulations)

Applicable from AY 2022-23 Batch

I Year I Semester

S. No.	Course Code	Course	L	T	P	Credits
1.	MA101BS	Matrices and Calculus	3	1	0	4
2.	CH102BS	Engineering Chemistry	3	1	0	4
3.	CS103ES	Programming for Problem Solving	3	0	0	3
4.	EE104ES	Basic Electrical Engineering	2	0	0	2
5.	ME105ES	Computer Aided Engineering Graphics	1	0	4	3
6.	CS106ES	Elements of Computer Science & Engineering	0	0	2	1
7.	CH107BS	Engineering Chemistry Laboratory	0	0	2	1
8.	CS108ES	Programming for Problem Solving Laboratory	0	0	2	1
9.	EE109ES	Basic Electrical Engineering Laboratory	0	0	2	1
		Induction Program				
		Total	**12**	**2**	**12**	**20**

I Year II Semester

S. No.	Course Code	Course	L	T	P	Credits
1.	MA201BS	Ordinary Differential Equations and Vector Calculus	3	1	0	4
2.	PH202BS	Applied Physics	3	1	0	4
3.	ME203ES	Engineering Workshop	0	1	3	2.5
4.	EN204HS	English for Skill Enhancement	2	0	0	2
5.	EC205ES	Electronic Devices and Circuits	2	0	0	2
6.	CS206ES	Python Programming Laboratory	0	1	2	2
7.	PH207BS	Applied Physics Laboratory	0	0	3	1.5
8.	EN208HS	English Language and Communication Skills Laboratory	0	0	2	1
9.	CS209ES	IT Workshop	0	0	2	1
10.	*MC210	Environmental Science	3	0	0	0
		Total	**13**	**4**	**12**	**20**

II YEAR I SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1	DS301PC	Digital Electronics	3	0	0	3
2	DS302PC	Data Structures	3	0	0	3
3	DS303PC	Computer Oriented Statistical Methods	3	1	0	4
4	DS304PC	Computer Organization and Architecture	3	0	0	3
5	DS305PC	Object Oriented Programming through Java	3	0	0	3
6	DS306PC	Data Structures Lab	0	0	3	1.5
7	DS307PC	Object Oriented Programming through Java Lab	0	0	3	1.5
8	DS308PC	Data visualization- R Programming/ Power BI	0	0	2	1
9	*MC309	Gender Sensitization Lab	0	0	2	0
		Total	**15**	**1**	**10**	**20**

II YEAR II SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1	DS401PC	Discrete Mathematics	3	0	0	3
2	SM402MS	Business Economics & Financial Analysis	3	0	0	3
3	DS403PC	Operating Systems	3	0	0	3
4	DS404PC	Database Management Systems	3	0	0	3
5	DS405PC	Software Engineering	3	0	0	3
6	DS406PC	Operating Systems Lab	0	0	2	1
7	DS407PC	Database Management Systems Lab	0	0	2	1
8	DS408PC	Real-time Research Project/ Societal Related Project	0	0	4	2
9	DS409PC	Node JS/ React JS/ Django	0	0	2	1
10	*MC410	Constitution of India	3	0	0	0
		Total	**18**	**0**	**10**	**20**

III YEAR I SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1	DS501PC	Algorithms Design and Analysis	3	0	0	3
2	DS502PC	Introduction to Data Science	3	1	0	4
3	DS503PC	Computer Networks	3	0	0	3
4		Professional Elective - I	3	0	0	3
5		Professional Elective - II	3	0	0	3
6	DS504PC	R Programming Lab	0	0	2	1
7	DS505PC	Computer Networks Lab	0	0	2	1
8	EN508HS	Advanced English Communication Skills Lab	0	0	2	1
9	DS506PC	ETL-Kafka/Talend	0	0	2	1
10	*MC510	Intellectual Property Rights	3	0	0	0
		Total	**18**	**1**	**08**	**20**

III YEAR II SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1	DS601PC	Automata Theory and Compiler Design	3	0	0	3
2	DS602PC	Machine Learning	3	0	0	3
3	DS603PC	Big Data Analytics	3	0	0	3
4		Professional Elective – III	3	0	0	3
5		Open Elective - I	3	0	0	3
6	DS604PC	Machine Learning Lab	0	0	2	1
7	DS605PC	Big Data Analytics Lab	0	0	2	1
8		Professional Elective - III Lab	0	0	2	1
9	DS606PC	Industrial Oriented Mini Project/ Summer Internship/ Skill Development Course (UI design- Flutter)	0	0	4	2
10	*MC609	Environmental Science	3	0	0	0
		Total	**18**	**0**	**10**	**20**

Environmental Science in III Yr II Sem Should be Registered by Lateral Entry Students Only.

IV YEAR I SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1	DS701PC	Predictive Analytics	3	0	0	3
2	DS702PC	Web and Social Media Analytics	3	0	0	3
3		Professional Elective – IV	3	0	0	3
4		Professional Elective – V	3	0	0	3
5		Open Elective – II	3	0	0	3
6	DS703PC	Predictive Analytics Lab	0	0	2	1
7	DS704PC	Web and Social Media Analytics Lab	0	0	2	1
8	DS705PC	Project Stage – I	0	0	6	3
		Total Credits	**15**	**0**	**10**	**20**

IV YEAR II SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1	DS801PC	Organizational Behavior	3	0	0	3
2		Professional Elective – VI	3	0	0	3
3		Open Elective – III	3	0	0	3
4	DS802PC	Project Stage – II including Seminar	0	0	22	11
		Total Credits	**9**	**0**	**22**	**20**

*MC – Satisfactory/Unsatisfactory

Professional Elective - I

DS511PE	Data Warehousing and Business Intelligence
DS512PE	Artificial Intelligence
DS513PE	Web Programming
DS514PE	Image Processing
DS515PE	Computer Graphics

Professional Elective - II

DS521PE	Spatial and Multimedia Databases
DS522PE	Information Retrieval Systems
DS523PE	Software Project Management
DS524PE	DevOps
DS525PE	Computer Vision and Robotics

Professional Elective - III

DS631PE	Software Testing Methodologies
DS632PE	Data Visualization Techniques
DS633PE	Scripting Languages
DS634PE	Mobile Application Development
DS635PE	Cryptography and Network Security

 # Courses in PE - III and PE - III Lab must be in 1-1 correspondence.

Professional Elective - IV

DS741PE	Quantum Computing
DS742PE	Database Security
DS743PE	Natural Language Processing
DS744PE	Information Storage Management
DS745PE	Internet of Things

Professional Elective - V

DS751PE	Privacy Preserving Data Publishing
DS752PE	Cloud Computing
DS753PE	Data Science Applications
DS754PE	Mining Massive Datasets
DS755PE	Exploratory Data Analysis

Professional Elective - VI

DS861PE	Data Stream Mining
DS862PE	Web Security
DS863PE	Video Analytics
DS864PE	Blockchain Technology
DS865PE	Parallel and Distributed Computing

Open Elective - I
1. DS611OE: Fundamentals of Data Science
2. DS612OE: R Programming

Open Elective - II
1. DS721OE: Data Mining
2. DS722OE: Data Analytics

Open Elective - III
1. DS831OE: Introduction to social media mining
2. DS832OE: Data Visualization using Python

DS701PC: PREDICTIVE ANALYTICS

B.Tech. IV Year I Sem.   **L   T   P   C**
                                                                                                                3   0   0   3

Prerequisite:
1. Data Science/Data analytics

Course Objectives:
- To learn the basics and applications of predictive analytics using different techniques

Course Outcomes:
- Understand the processing steps for predictive analytics
- Construct and deploy prediction models with integrity
- Explore various techniques (machine learning/data mining, ensemble) for predictive analytics.
- Apply predictive analytics to real world examples.

UNIT - I
Introduction – types of analytics, applications of predictive analytics, overview of predictive analytics. Setting up the problem - processing steps, business understanding, objectives, data for predictive modeling, columns as measures, target variables, measures of success for predictive models.

UNIT - II
Prediction effect, deployment of prediction model, ethics and responsibilities The Data effect

UNIT - III
Machine Learning for prediction
Predictive modeling – decision trees, logistic regression, neural network, kNN, Bayesian method,
Regression model
Assessing Predictive models - Batch Approach to Model Assessment, Percent Correct Classification, Rank-Ordered Approach to Model Assessment, Assessing Regression Models

UNIT - IV
Ensemble effect
Model ensembles – motivation, wisdom of crowds, Bagging, Boosting, Random forests, stochastic gradient boosting, heterogeneous ensembles.

UNIT - V
Case studies: Survey analysis, question answering– challenges in text mining, persuasion by the numbers

TEXT BOOKS:
1. Eric Siegel, Predictive analytics- the power to predict who will Click, buy, lie, or die, John Wiley & Sons, 2013.
2. Dean Abbott, Applied Predictive Analytics - Principles and Techniques for the Professional Data Analyst, 2014.

REFERENCE BOOKS:
1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009.
2. G. James, D. Witten, T. Hastie, R. Tibshirani-An introduction to statistical learning with applications in R, Springer, 2013.
3. E. Alpaydin, Introduction to Machine Learning, Prentice Hall of India, 2010.

DS702PC: WEB AND SOCIAL MEDIA ANALYTICS

B.Tech. IV Year I Sem.	L	T	P	C
	3	0	0	3

Course Objectives:
- Exposure to various web and social media analytic techniques.

Course Outcomes:
- Knowledge on decision support systems
- Apply natural language processing concepts on text analytics
- Understand sentiment analysis
- Knowledge on search engine optimization and web analytics

UNIT - I
An Overview of Business Intelligence, Analytics, and Decision Support
Analytics to Manage a Vaccine Supply Chain Effectively and Safely, Changing Business Environments and Computerized Decision Support, Information Systems Support for Decision Making, The Concept of Decision Support Systems (DSS), Business Analytics Overview, Brief Introduction to Big Data Analytics

UNIT - II
Text Analytics and Text Mining
Machine Versus Men on Jeopardy: The Story of Watson, Text Analytics and Text Mining Concepts and Definitions, Natural Language Processing, Text Mining Applications, Text Mining Process, Text Mining Tools

UNIT - III
Sentiment Analysis
Sentiment Analysis Overview, Sentiment Analysis Applications, Sentiment Analysis Process, Sentiment Analysis and Speech Analytics

UNIT - IV
Web Analytics, Web Mining
Security First Insurance Deepens Connection with Policyholders, Web Mining Overview, Web Content and Web Structure Mining, Search Engines, Search Engine Optimization, Web Usage Mining (Web Analytics), Web Analytics Maturity Model and Web Analytics Tools

UNIT - V
Social Analytics and Social Network Analysis
Social Analytics and Social Network Analysis, Social Media Definitions and Concepts, Social Media Analytics
Prescriptive Analytics - Optimization and Multi-Criteria Systems:
Multiple Goals, Sensitivity Analysis, What-If Analysis, and Goal Seeking

TEXT BOOK:
1. Ramesh Sharda, Dursun Delen, Efraim Turban, Business Intelligence and Analytics: Systems for Decision Support, Pearson Education
REFERENCE BOOKS:
1. Rajiv Sabherwal, Irma Becerra- Fernandez," Business Intelligence– Practice, Technologies and Management", John Wiley 2011.
2. Lariss T. Moss, ShakuAtre, "Business Intelligence Roadmap", Addison-Wesley It Service.
3. Yuli Vasiliev, "Oracle Business Intelligence: The Condensed Guide to Analysis and Reporting", SPD Shroff, 2012.

DS741PE: QUANTUM COMPUTING (Professional Elective – IV)

B.Tech. IV Year I Sem.	L	T	P	C
	3	0	0	3

Course Objectives
- To introduce the fundamentals of quantum computing
- The problem-solving approach using finite dimensional mathematics

Course Outcomes
- Understand basics of quantum computing
- Understand physical implementation of Qubit
- Understand Quantum algorithms and their implementation
- Understand The Impact of Quantum Computing on Cryptography

UNIT - I
History of Quantum Computing: Importance of Mathematics, Physics and Biology. Introduction to Quantum Computing: Bits Vs Qubits, Classical Vs Quantum logical operations

UNIT - II
Background Mathematics: Basics of Linear Algebra, Hilbert space, Probabilities and measurements. Background Physics: Paul's exclusion Principle, Superposition, Entanglement and super-symmetry, density operators and correlation, basics of quantum mechanics, Measurements in bases other than computational basis. Background Biology: Basic concepts of Genomics and Proteomics (Central Dogma)

UNIT - III
Qubit: Physical implementations of Qubit. Qubit as a quantum unit of information. The Bloch sphere Quantum Circuits: single qubit gates, multiple qubit gates, designing the quantum circuits. Bell states.

UNIT - IV
Quantum Algorithms: Classical computation on quantum computers. Relationship between quantum and classical complexity classes. Deutsch's algorithm, Deutsch's-Jozsa algorithm, Shor's factorization algorithm, Grover's search algorithm.

UNIT - V
Noise and error correction: Graph states and codes, Quantum error correction, fault-tolerant computation. Quantum Information and Cryptography: Comparison between classical and quantum information theory. Quantum Cryptography, Quantum teleportation

TEXT BOOK:
1. Nielsen M. A., Quantum Computation and Quantum Information, Cambridge.

REFERENCE BOOKS:
1. Quantum Computing for Computer Scientists by Noson S. Yanofsky and Mirco A. Mannucci
2. Benenti G., Casati G. and Strini G., Principles of Quantum Computation and Information, Vol. I: Basic Concepts, Vol II.
3. Basic Tools and Special Topics, World Scientific. Pittenger A. O., An Introduction to Quantum Computing Algorithms.

DS742PE: DATABASE SECURITY (Professional Elective – IV)

B.Tech. IV Year I Sem.	L	T	P	C
	3	0	0	3

Course Objectives:
- To learn the security of databases
- To learn the design techniques of database security
- To learn the security software design

Course Outcomes:
- Identify database security problems
- Implement different security models
- Provide security for software design
- Protect object-oriented systems
- Handle security issues for active databases

UNIT - I
Introduction: Introduction to Databases Security, Problems in Databases Security, Controls, Conclusions
Security Models -1: Introduction Access Matrix Model, Take-Grant Model, Acten Model, PN Model.

UNIT - II
Security Models -2: Hartson and Hsiao's Model, Fernandez's Model, Bussolati and Martella's Model for Distributed databases, Bell and LaPadula's Model, Biba's Model, Dion's Model, Sea View Model, Jajodia and Sandhu's Model, The Lattice Model for the Flow Control.

UNIT - III
Security Mechanisms: Introduction User Identification/Authentication, Memory Protection, Resource Protection, Control Flow Mechanisms, Isolation Security Functionalities in Some Operating Systems Trusted Computer System Evaluation Criteria

Security Software Design: Introduction A Methodological Approach to Security Software Design Secure Operating System Design Secure DBMS Design Security Packages Database Security Design

UNIT - IV
Statistical Database Protection & **Intrusion Detection Systems:** Introduction Statistics Concepts and Definitions Types of Attacks Inference Controls Evaluation Criteria for Control Comparison. Introduction IDES System RETISS System ASES System Discovery

UNIT - V
Models For the Protection of New Generation Database Systems: Introduction A Model for the Protection of Frame Based Systems A Model for the Protection of Object-Oriented Systems SORION Model for the Protection of Object-Oriented Databases, The Orion Model, Jajodia and Kogan's Model-A Model for the Protection of Active Databases Conclusions

TEXT BOOKS:
1. Database Security by Castano *Pearson Edition (*1/e)

REFERENCE BOOK:
1. Database security by alfred basta, melissa zgola, CENGAGE learning.
2. Database Security and Auditing: Protecting Data Integrity and Accessibility, 1st Edition, Hassan Afyouni, THOMSON Edition.

DS743PE: NATURAL LANGUAGE PROCESSING (Professional Elective – IV)

B.Tech. IV Year I Sem.   **L T P C**
                                                                    3 0 0 3

Prerequisites:
- Data structures and compiler design

Course Objectives:
- Introduction to some of the problems and solutions of NLP and their relation to linguistics and statistics.

Course Outcomes:
- Show sensitivity to linguistic phenomena and an ability to model them with formal grammars.
- Understand and carry out proper experimental methodology for training and evaluating empirical NLP systems
- Able to manipulate probabilities, construct statistical models over strings and trees, and estimate parameters using supervised and unsupervised training methods.
- Able to design, implement, and analyze NLP algorithms; and design different language modeling Techniques.

UNIT - I
Finding the Structure of Words: Words and Their Components, Issues and Challenges, Morphological Models
Finding the Structure of Documents: Introduction, Methods, Complexity of the Approaches, Performances of the Approaches, Features

UNIT - II
Syntax I: Parsing Natural Language, Treebanks: A Data-Driven Approach to Syntax, Representation of Syntactic Structure, Parsing Algorithms

UNIT – III
Syntax II: Models for Ambiguity Resolution in Parsing, Multilingual Issues
Semantic Parsing I: Introduction, Semantic Interpretation, System Paradigms, Word Sense

UNIT - IV
Semantic Parsing II: Predicate-Argument Structure, Meaning Representation Systems

UNIT - V
Language Modeling: Introduction, N-Gram Models, Language Model Evaluation, Bayesian parameter estimation, Language Model Adaptation, Language Models- class based, variable length, Bayesian topic based, Multilingual and Cross Lingual Language Modeling

TEXT BOOKS:
1. Multilingual natural Language Processing Applications: From Theory to Practice – Daniel M. Bikel and Imed Zitouni, Pearson Publication

REFERENCE BOOK:
1. Speech and Natural Language Processing - Daniel Jurafsky& James H Martin, Pearson Publications.
2. Natural Language Processing and Information Retrieval: Tanvier Siddiqui, U.S. Tiwary.

DS744PE: INFORMATION STORAGE MANAGEMENT (Professional Elective – IV)

B.Tech. IV Year I Sem.　　　　　　　　　　　　　　　　　　　　　**L　T　P　C**
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　3　0　0　3

Course Objectives:
- To understand various segments of storage technology and architectures
- To explore the inherent power of information
- To describe the different backup, recovery and replication strategies

Course Outcomes:
- Understand the evolution of storage technology and Intelligent Storage Systems
- Explore the key concepts of various Storage Networking Technologies - DAS, SANs, NAS and CAS
- Understand the basics of Storage Virtualization
- Understand the concepts of Storage security and Storage Infrastructure Management
- Analyze the purpose of backup, recovery and replication Strategies

UNIT - I
Introduction to Information Storage and Management: Information Storage, Evolution of Storage Technology and Architecture, Data Center Infrastructure, Key Challenges in Managing Information, Information Lifecycle. Storage System Environment - Data Protection: RAID - Intelligent Storage System.

UNIT - II
Direct-Attached Storage and Introduction to SCSI
Types of DAS, DAS Benefits and Limitations, Disk Drive Interfaces, Introduction to Parallel SCSI, SCSI Command Model,
Storage Area Networks
Fibre Channel: Overview, The SAN and Its Evolution, Components of SAN, FC Connectivity, Fibre Channel Ports, Fibre Channel Architecture, Zoning, Fibre Channel Login Types, FC Topologies, Concepts in Practice: EMC Connectrix
Network-Attached Storage
General-Purpose Servers vs. NAS Devices, Benefits of NAS, NAS File I/O, Components of NAS, NAS Implementations, NAS File-Sharing Protocols, NAS I/O Operations, Factors Affecting NAS Performance and Availability, Concepts in Practice: EMC Celerra

UNIT - III
Content-Addressed Storage
Fixed Content and Archives, Types of Archives, Features and Benefits of CAS, CAS Architecture, Object Storage and Retrieval in CAS, CAS Examples, Concepts in Practice: EMC Centera
Storage Virtualization
Forms of Virtualization, SNIA Storage Virtualization Taxonomy, Storage Virtualization Configurations, Storage Virtualization Challenges, Types of Storage Virtualization, Concepts in Practice

UNIT - IV
Backup and Recovery
Backup Purpose, Backup Considerations, Backup Granularity, Recovery Considerations, Backup Methods, Backup Process, Backup and Restore Operations, Backup Topologies, Backup in NAS Environments, Backup Technologies, Concepts in Practice: EMC NetWorker
Local Replication

Local Replication, Source and Target, Uses of Local Replicas, Data Consistency, Local Replication Technologies, Restore and Restart Considerations, Creating Multiple Replicas, Management Interface, Concepts in Practice: EMC TimeFinder and EMC SnapView

Remote Replication

Modes of Remote Replication, Remote Replication Technologies, Network Infrastructure, Concepts in Practice: EMC SRDF, EMC SAN Copy, and EMC MirrorView

UNIT - V
Securing the Storage Infrastructure
Storage Security Framework, Risk Triad, Storage Security Domains, Security Implementations in Storage Networking

Managing the Storage Infrastructure
Monitoring the Storage Infrastructure, Storage Management Activities, Storage Infrastructure Management Challenges, Developing an Ideal Solution, Concepts in Practice: EMC ControlCenter

TEXT BOOKS:
1. Marc Farley Osborne, "Building Storage Networks", Tata McGraw Hill, 2001.
2. Robert Spalding and Robert Spalding, "Storage Networks: The Complete Reference", Tata McGraw Hill, 2003.
3. Meeta Gupta, "Storage Area Network Fundamentals", Pearson Education Ltd., 2002.

REFERENCE BOOKS:
1. Gerald J Kowalski and Mark T Maybury," Information Storage Retrieval Systems theory & Implementation", BS Publications, 2000.
2. Thejendra BS, "Disaster Recovery & Business continuity", Shroff Publishers & Distributors, 2006.

DS745PE: INTERNET OF THINGS (Professional Elective – IV)

B.Tech. IV Year I Sem.   **L T P C**
                                                                               3 0 0 3

Pre-Requisites: Computer organization, Computer Networks
Course Objectives:
- To introduce the terminology, technology and its applications
- To introduce the concept of M2M (machine to machine) with necessary protocols
- To introduce the Python Scripting Language which is used in many IoT devices
- To introduce the Raspberry PI platform, that is widely used in IoT applications
- To introduce the implementation of web-based services on IoT devices

Course Outcomes:
- Interpret the impact and challenges posed by IoT networks leading to new architectural models.
- Compare and contrast the deployment of smart objects and the technologies to connect them to network.
- Appraise the role of IoT protocols for efficient network communication.
- Identify the applications of IoT in Industry.

UNIT - I
Introduction to Internet of Things –Definition and Characteristics of IoT, Physical Design of IoT, Logical Design of IoT, IoT Enabling Technologies, IoT Levels and Deployment Templates
Domain Specific IoTs – Home automation, Environment, Agriculture, Health and Lifestyle

UNIT - II
IoT and M2M – M2M, Difference between IoT and M2M, SDN and NFV for IoT,
IoT System Management with NETCOZF, YANG- Need for IoT system Management, Simple Network management protocol, Network operator requirements, NETCONF, YANG, IoT Systems Management with NETCONF-YANG

UNIT - III
IoT Systems – Logical design using Python-Introduction to Python – Python Data types & Data structures, Control flow, Functions, Modules, Packaging, File handling, Data/Time operations, Classes, Exception, Python packages of Interest for IoT

UNIT - IV
IoT Physical Devices and Endpoints - Raspberry Pi, Linux on Raspberry Pi, Raspberry Pi Interfaces, Programming Raspberry PI with Python, Other IoT devices.

IoT Physical Servers and Cloud Offerings – Introduction to Cloud Storage models and communication APIs, WAMP-AutoBahn for IoT, Xively Cloud for IoT, Python web application framework –Django, Designing a RESTful web API

UNIT V
Case studies- Home Automation, Environment-weather monitoring-weather reporting- air pollution monitoring, Agriculture.

TEXT BOOK:
1. Internet of Things - A Hands-on Approach, Arshdeep Bahga and Vijay Madisetti, Universities Press, 2015, ISBN: 9788173719547.

REFERENCE BOOK:
1. Getting Started with Raspberry Pi, Matt Richardson & Shawn Wallace, O'Reilly (SPD), 2014, ISBN: 9789350239759.

DS751PE: PRIVACY PRESERVING DATA PUBLISHING (Professional Elective – V)

B.Tech. IV Year I Sem.   L T P C
                                                                                              3 0 0 3

Prerequisites
- A course on "Data Mining".

Course Objectives
- The aim of the course is to introduce the fundamentals of Privacy Preserving Data Mining Methods
- The course gives an overview of - Anonymity and its Measures, Multiplicative Perturbation for Privacy-Preserving Data Mining, techniques for Utility-based Privacy Preserving Data

Course Outcomes
- Understand the concepts of Privacy Preserving Data Mining Models and Algorithms
- Demonstrate a comprehensive understanding of different tasks associated in Inference Control Methods for Privacy-Preserving Data Mining
- Understand the concepts of Data Anonymization Methods and its Measures
- Evaluate and Appraise the solution designed for Multiplicative Perturbation
- Formulate, Design and Implement the solutions for Utility-based Privacy Preserving Data

UNIT-I
Introduction, Privacy-Preserving Data Mining Algorithms, The Randomization Method, Group Based Anonymization, Distributed Privacy-Preserving Data Mining

UNIT -II
Interface Control Methods
Introduction, A Classification of Microdata Protection Methods, Perturbative Masking Methods, Non-Perturbative Masking Methods, Synthetic Microdata Generation, Trading off Information Loss and Disclosure Risk.

UNIT -III
Measure of Anonymity
Data Anonymization Methods, A Classification of Methods, Statistical Measure of Anonymous, Probabilistic Measure of Anonymity, Computational Measure of Anonymity, reconstruction Methods for Randomization, Application of Randomization

UNIT-IV
Multiplicative Perturbation
Definition of Multiplicative Perturbation, Transformation Invariant Data Mining Models, Privacy Evaluation for Multiplicative Perturbation, Attack Resilient Multiplicative Perturbation, Metrics for Quantifying Privacy Level, Metrics for Quantifying Hiding Failure, Metrics for Quantifying Data Quality.

UNIT-V
Utility-Based Privacy-Preserving Data
Types of Utility-Based Privacy Preserving Methods, Utility-Based Anonymization Using Local Recording, The Utility-Based Privacy Preserving Methods in Classification Problems, Anonymization Merginal: Injection Utility into Anonymization Data Sets.

Text Book:
1. Privacy – Preserving Data Mining: Models and Algorithms Edited by Charu C. Aggarwal and S. Yu, Springer

Reference Books:
1. Charu C. Agarwal, Data Mining: The Textbook, 1st Edition, Springer.
2. Han and M. Kamber, Data Mining: Concepts and Techniques, 3rd Edition, Elsevier.
3. Privacy Preserving Data Mining by Jaideep Vaidya, Yu Michael Zhu and Chirstopher W. Clifton, Springer

DS752PE: CLOUD COMPUTING (Professional Elective – V)

B.Tech. IV Year I Sem.   **L   T   P   C**
                                                                        3   0   0   3

Pre-requisites:
1. A course on "Computer Networks".
2. A course on "Operating System".

Course Objectives:
- This course provides an insight into cloud computing.
- Topics covered include- Cloud Computing Architecture, Deployment Models, Service Models, Technological Drivers for Cloud Computing, Networking for Cloud Computing and Security in Cloud Computing.

Course Outcomes:
- Understand different computing paradigms and potential of the paradigms and specifically cloud computing
- Understand cloud service types, cloud deployment models and technologies supporting and driving the cloud
- Acquire the knowledge of programming models for cloud and development of software application that runs the cloud and various services available from major cloud providers
- Understand the security concerns and issues in cloud computing
- Acquire the knowledge of advances in cloud computing.

UNIT - I
Computing Paradigms, Cloud Computing Fundamentals, Cloud Computing Architecture and Management

UNIT - II
Cloud Deployment Models, Cloud Service Models, Technological Drivers for Cloud Computing: SOA and Cloud, Multicore Technology, Web 2.0 and Web 3.0, Pervasive Computing, Operating System, Application Environment

UNIT - III
Virtualization, Programming Models for Cloud Computing: MapReduce, Cloud Haskell, Software Development in Cloud

UNIT - IV
Networking for Cloud Computing: Introduction, Overview of Data Center Environment, Networking Issues in Data Centers, Transport Layer Issues in DCNs, Cloud Service Providers

UNIT - V
Security in Cloud Computing, and Advanced Concepts in Cloud Computing

TEXT BOOK:
1. Chandrasekaran, K. *Essentials of cloud computing*. CRC Press, 2014
REFERENCE BOOKS:
1. Cloud Computing: Principles and Paradigms, Editors: Rajkumar Buyya, James Broberg, Andrzej M. Goscinski, Wiley, 2011
2. Enterprise Cloud Computing - Technology, Architecture, Applications, Gautam Shroff, Cambridge University Press, 2010
3. Cloud Computing Bible, Barrie Sosinsky, Wiley-India, 2010

DS753PE: DATA SCIENCE APPLICATIONS (Professional Elective – V)

B.Tech. IV Year I Sem.		L	T	P	C
		3	0	0	3

Course Objective:
- To give deep knowledge of data science and how it can be applied in various fields to make life easy.

Course Outcomes:
- Correlate data science and solutions to modern problems.
- Decide when to use which type of technique in data science.

UNIT - I
Data Science Applications in various domains, Challenges and opportunities, tools for data scientists
Recommender systems – Introduction, methods, application, challenges.

UNIT - II
Time series data – stock market index movement forecasting.
Supply Chain Management – Real world case study in logistics

UNIT - III
Data Science in Education, social media

UNIT - IV
Data Science in Healthcare, Bioinformatics

UNIT - V
Case studies in data optimization using Python.

TEXT BOOKS:
1. Aakanksha Sharaff, G.K. Sinha, "Data Science and its applications ", CRC Press, 2021.
2. Q.A. Menon, S.A. Khoja, "Data Science: Theory, Analysis and Applications", CRC Press, 2020

DS754PE: MINING MASSIVE DATASETS (Professional Elective – V)

B.Tech. IV Year I Sem.   **L T P C**
                                                                                        3 0 0 3

Prerequisites:
1. Students should be familiar with Data mining, algorithms, basic probability theory and Discrete math.

Course Objectives:
1. This course will cover practical algorithms for solving key problems in mining of massive datasets.
2. This course focuses on parallel algorithmic techniques that are used for large datasets.
3. This course will cover stream processing algorithms for data streams that arrive constantly, page ranking algorithms for web search, and online advertisement systems that are studied in detail.

Course Outcomes:
1. Handle massive data using MapReduce.
2. Develop and implement algorithms for massive data sets and methodologies in the   context of data mining.
3. Understand the algorithms for extracting models and information from large datasets
4. Develop recommendation systems.
5. Gain experience in matching various algorithms for particular classes of problems.

UNIT - I
Data Mining-Introduction-Definition of Data Mining-Statistical Limits on Data Mining,
MapReduce and the New Software Stack-Distributed File Systems, MapReduce, Algorithms Using MapReduce.

UNIT - II
Similarity Search: Finding Similar Items-Applications of Near-Neighbor Search, Shingling of Documents, Similarity-Preserving Summaries of Sets, Distance Measures.
Streaming Data: Mining Data Streams-The Stream Data Model , Sampling Data in a Stream, Filtering Streams

UNIT - III
Link Analysis-PageRank, Efficient Computation of PageRank, Link Spam
Frequent Itemsets-Handling Larger Datasets in Main Memory, Limited-Pass Algorithms, Counting Frequent Items in a Stream.
Clustering-The CURE Algorithm, Clustering in Non-Euclidean Spaces, Clustering for Streams and Parallelism

UNIT - IV
Advertising on the Web-Issues in On-Line Advertising, On-Line Algorithms, The Matching Problem, The Adwords Problem, Adwords Implementation.
Recommendation Systems-A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering, Dimensionality Reduction, The NetFlix Challenge.

UNIT - V:
Mining Social-Network Graphs-Social Networks as Graphs, Clustering of Social-Network Graphs, Partitioning of Graphs, Simrank, Counting Triangles

TEXT BOOKS:
1. Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining of Massive Datasets, 3rd Edition.

Reference Books:
1. Jiawei Han & Micheline Kamber, Data Mining – Concepts and Techniques 3rd Edition Elsevier.
2. Margaret H Dunham, Data Mining Introductory and Advanced topics, PEA.
3. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann.

DS755PE: EXPLORATORY DATA ANALYSIS (Professional Elective – V)

B.Tech. IV Year I Sem.   **L   T   P   C**
                                                                                 3   0   0   3

Course Objectives:
- Analysis of data, exploring various models in exploratory data analysis, question answering and predictive analysis

Course Outcomes:
- Apply the Epicycle of Analysis process effectively.
- Articulate and refine data-related questions using the Epicycle approach.
- Conduct Exploratory Data Analysis (EDA).
- Develop the skills necessary to use formal modeling techniques for data inference.

UNIT – I
Epicycles of Analysis: Setting the Scene, Epicycle of Analysis, Setting Expectations, Collecting Information, Comparing Expectations to Data, Applying the Epicycle of Analysis process.

UNIT – II
Stating and Refining the Question: Types of Questions, Applying the Epicycle to stating and Refining Your Question, Characteristics of good Question, Translating a Question into a Data Problem, Case Study.

UNIT- III
Exploratory Data Analysis: Formulate your question, read in your data, Checking Packaging, look at the top and bottom of the data, always be checking, validate with at least one External Source, make a plot, Try the Easy Solution First.

UNIT – IV
Using Models to Explore your data: Models as Expectations, Reacting to Data Refining Our Expectations, Examining Linear Relationships, Stopping Criteria.
Inference: Identify the population, Describe the sampling process, Describe the Model for the population, Factors Affecting the Quality of Inference, Case Study.

UNIT – V
Formal Modeling: Goals of Formal Modeling, General Frame work, Associational Analysis, Prediction Analysis, and Summary

TEXT BOOK:
1. "The Art of Data Science: A Guide for Anyone Who Works with Data" by Roger D. Peng and Elizabeth Matsui.

REFERENCE BOOKS:
1. "Exploratory Data Analytics "by John Tukey.
2. "Python for Data Analysis "by Wes McKinney

DS721OE: DATA MINING (Open Elective – II)

B.Tech. IV Year I Sem.	L T P C
	3 0 0 3

Pre-Requisites:
1. Database Management System
2. Probability and Statistics

Course Objectives:
- Students will become acquainted with both the strengths and limitations of various data mining techniques like Association, Classification, Cluster and Outlier analysis.

Course Outcomes:
- Understand the need of data mining and pre-processing techniques.
- Perform market basket analysis using association rule mining.
- Utilize classification techniques for analysis and interpretation of data.
- Identify appropriate clustering and outlier detection techniques to handle complex data.
- Understand the mining of data from web, text and time series data.

UNIT - I
Introduction to Data Mining:
What Data mining? Kinds of Data, Knowledge Discovery process, Data Mining Functionalities, Kinds of Patterns, Major Issues in Data Mining. Data Objects and Attribute Types, Basic Statistical Descriptions of Data, Data Visualization, Measuring Data Similarity and Dissimilarity, Data Pre-processing: Major Tasks in Data Pre-processing, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization.

UNIT - II
Association Analysis: Basic Concepts, Market Basket Analysis, Apriori Algorithm, FP-growth, From Association Analysis to Correlation Analysis, Pattern Mining in Multilevel Associations and Multidimensional Associations.

UNIT - III
Classification: Basic Concepts, Decision Tree Induction, Bayes Classification Methods, Rule-Based Classification, Metrics for Evaluating Classifier Performance, Ensemble Methods, Multilayer Feed-Forward Neural Network, Support Vector Machines, k-Nearest-Neighbor Classifiers.

UNIT - IV
Cluster Analysis: Requirements for Cluster Analysis, Overview of Basic Clustering Methods, Partitioning Methods-k-Means, k-Medoids, Hierarchical Methods-AGENES, DIANA, BIRCH, Density-Based Method-DBSCAN, Outlier Analysis: Types of Outliers, Challenges of Outlier Detection, and Overview of Outlier Detection Methods

UNIT - V
Advanced Concepts: Web Mining- Web Content Mining, Web Structure Mining, Web Usage Mining, Spatial Mining- Spatial Data Overview, Spatial Data Mining Primitives, Spatial Rules, Spatial Classification Algorithm, Spatial Clustering Algorithms, Temporal Mining- Modeling Temporal Events, Time Series, Pattern Detection, Sequences, Temporal Association Rules.

TEXT BOOKS:
1. Jiawei Han, Micheline Kamber, Jian Pei., Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann/Elsevier, 2012.
2. Margaret H Dunham, Data Mining Introductory and Advanced Topics, 2nd Edition, Pearson Education, India, 2006.

REFERENCE BOOKS:
1. Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
2. Pang-Ning Tan, Michael Steinbach, Anuj Karpatne and Vipin Kumar, Introduction to Data Mining, 2nd Edition, Pearson Education India, 2021.
3. Amitesh Sinha, Data Warehousing, Thomson Learning, India, 2007.

DS722OE: DATA ANALYTICS (Open Elective – II)

B.Tech. IV Year I Sem.	**L**	**T**	**P**	**C**
	3	**0**	**0**	**3**

Prerequisites
1. A course on "Database Management Systems".
2. Knowledge of probability and statistics.

Course Objectives:
- To explore the fundamental concepts of data analytics.
- To learn the principles and methods of statistical analysis
- Discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms.
- To understand the various search methods and visualization techniques.

Course Outcomes: After completion of this course students will be able to
- Understand the impact of data analytics for business decisions and strategy
- Carry out data analysis/statistical analysis
- To carry out standard data visualization and formal inference procedures
- Design Data Architecture
- Understand various Data Sources

UNIT - I
Data Management: Design Data Architecture and manage the data for analysis, understand various sources of Data like Sensors/Signals/GPS etc. Data Management, Data Quality(noise, outliers, missing values, duplicate data) and Data Processing & Processing.

UNIT - II
Data Analytics: Introduction to Analytics, Introduction to Tools and Environment, Application of Modeling in Business, Databases & Types of Data and Variables, Data Modeling Techniques, Missing Imputations etc. Need for Business Modeling.

UNIT - III
Regression – Concepts, Blue property assumptions, Least Square Estimation, Variable Rationalization, and Model Building etc.
Logistic Regression: Model Theory, Model fit Statistics, Model Construction, Analytics applications to various Business Domains etc.

UNIT - IV
Object Segmentation: Regression Vs Segmentation – Supervised and Unsupervised Learning, Tree Building – Regression, Classification, Overfitting, Pruning and Complexity, Multiple Decision Trees etc.
Time Series Methods: Arima, Measures of Forecast Accuracy, STL approach, Extract features from generated model as Height, Average Energy etc and Analyze for prediction

UNIT - V
Data Visualization: Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data and Relations.

TEXT BOOKS:
1. Student's Handbook for Associate Analytics – II, III.

2. Data Mining Concepts and Techniques, Han, Kamber, 3rd Edition, Morgan Kaufmann Publishers.

REFERENCE BOOKS:
1. Introduction to Data Mining, Tan, Steinbach and Kumar, Addision Wisley, 2006.
2. Data Mining Analysis and Concepts, M. Zaki and W. Meira
3. Mining of Massive Datasets, Jure Leskovec Stanford Univ. Anand Rajaraman Milliway Labs Jeffrey D Ullman Stanford Univ.

DS703PC: PREDICTIVE ANALYTICS LAB

B.Tech. IV Year I Sem.   **L   T   P   C**
                                                                     0   0   2   1

Course Objectives:
- To learn the basics and applications of predictive analytics using different techniques

Course Outcomes:
- Understand the processing steps for predictive analytics
- Construct and deploy prediction models with integrity
- Explore various techniques (machine learning/data mining, ensemble) for predictive analytics.
- Apply predictive analytics to real world examples.

List of Experiments: Following experiments to be carried out using Python/SPSS/SAS/R/Power BI
1. Simple Linear regression
2. Multiple Linear regression
3. Logistic Regression
4. CHAID
5. CART
6. ARIMA – stock market data
7. Exponential Smoothing
8. Hierarchical clustering
9. Ward's method of clustering
10. Crowdsource predictive analytics- Netflix data

TEXT BOOKS:
1. Eric Siegel, Predictive analytics- the power to predict who will Click, buy, lie, or die, John Wiley & Sons, 2013.
2. Dean Abbott, Applied Predictive Analytics - Principles and Techniques for the Professional Data Analyst, 2014.

REFERENCE BOOKS:
1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009.
2. G. James, D. Witten, T. Hastie, R. Tibshirani-An introduction to statistical learning with applications in R, Springer, 2013
3. E. Alpaydin, Introduction to Machine Learning, Prentice Hall of India, 2010

DS704PC: WEB AND SOCIAL MEDIA ANALYTICS LAB

B.Tech. IV Year I Sem.   **L T P C**
                                                                              0 0 2 1

Course Objectives:
- Exposure to various web and social media analytic techniques.

Course Outcomes:
- Knowledge on decision support systems
- Apply natural language processing concepts on text analytics
- Understand sentiment analysis
- Knowledge on search engine optimization and web analytics

List of Experiments
1. Preprocessing text document using NLTK of Python
   a. Stopword elimination
   b. Stemming
   c. Lemmatization
   d. POS tagging
   e. Lexical analysis
2. Sentiment analysis on customer review on products
3. Web analytics
   a. Web usage data (web server log data, clickstream analysis)
   b. Hyperlink data
4. Search engine optimization- implement spamdexing
5. Use Google analytics tools to implement the following
   a. Conversion Statistics
   b. Visitor Profiles
6. Use Google analytics tools to implement the Traffic Sources.

Resources:
1. Stanford core NLP package
2. GOOGLE.COM/Analytics

TEXT BOOK:
1. Ramesh Sharda, Dursun Delen, Efraim Turban, Business Intelligence and Analytics: Systems for Decision Support, Pearson Education

REFERENCE BOOKS:
1. Rajiv   Sabherwal, Irma Becerra-Fernandez, "Business Intelligence–Practice, Technologies and Management", John Wiley 2011.
2. Lariss T. Moss, Shaku Atre, "Business Intelligence Roadmap", Addison-Wesley It Service.
3. Yuli Vasiliev, "Oracle Business Intelligence: The Condensed Guide to Analysis and Reporting", SPD Shroff, 2012.

DS801PC: ORGANIZATIONAL BEHAVIOUR

B.Tech. IV Year II Sem.   **L T P C**
                          3 0 0 3

Course Objectives:
- This course demonstrates individual, group behavior aspects: The dynamics of organizational climate, structure and its impact on Organizations.

Course Outcomes:
- Students understand their personality, perception and attitudes for overall development and further learn the importance of group behavior in the organizations.

UNIT - I Organizational Behaviour
Definition, need and importance of organizational behaviour – Nature and scope – Frame work – Organizational behaviour models.

UNIT - II Individual Behaviour
Personality – types – Factors influencing personality – Theories – Learning – Types of learners – The learning process – Learning theories – Organizational behaviour modification, Misbehaviour – Types – Management Intervention. Emotions - Emotional Labour – Emotional Intelligence – Theories. Attitudes – Characteristics – Components – Formation – Measurement- Values. Perceptions – Importance – Factors influencing perception – Interpersonal perception- Impression Management. Motivation – importance – Types – Effects on work behavior.

UNIT - III Group Behaviour
Organization structure – Formation – Groups in organizations – Influence – Group dynamics – Emergence of informal leaders and working norms – Group decision making techniques – Team building - Interpersonal relations – Communication – Control.

UNIT - IV Leadership and Power
Meaning – Importance – Leadership styles – Theories of leadership – Leaders Vs Managers – Sources of power – Power centers – Power and Politics.

UNIT - V Dynamics of Organizational Behaviour
Organizational culture and climate – Factors affecting organizational climate – Importance. Job satisfaction – Determinants – Measurements – Influence on behavior. Organizational change – Importance – Stability Vs Change – Proactive Vs Reaction change – the change process – Resistance to change – Managing change. Stress – Work Stressors – Prevention and Management of stress – Balancing work and Life. Organizational development – Characteristics – objectives –. Organizational effectiveness

TEXT BOOKS:
1. Stephen P. Robins, Organisational Behavior, PHI Learning / Pearson Education, 11th edition, 2008.
2. Fred Luthans, Organisational Behavior, McGraw Hill, 11th Edition, 2001.

REFERENCE BOOKS:
1. Schermerhorn, Hunt and Osborn, Organisational behavior, John Wiley, 9th Edition, 2008.
2. Udai Pareek, Understanding Organisational Behaviour, 2nd Edition, Oxford Higher Education, 2004.

DS861PE: DATA STREAM MINING (Professional Elective – VI)

B.Tech. IV Year I Sem.	**L**	**T**	**P**	**C**
	3	**0**	**0**	**3**

Prerequisites
- A basic knowledge of "Data Mining"

Course Objectives
- The aim of the course is to introduce the fundamentals of Data Stream Mining.
- The course gives an overview of – Mining Strategies, methods and algorithms for data stream mining.

Course Outcomes
- Understand how to formulate a knowledge extraction problem from data streams.
- Ability to apply methods / algorithms to new data stream analysis problems.
- Evaluate the results and understand the functioning of the methods studied.
- Demonstrate decision tree and adaptive Hoeffding Tree concepts

UNIT- I
MOA Stream Mining, Assumptions, Requirements, Mining Strategies, Change Detection Strategies, MOA Experimental Settings, Previous Evaluation Practices, Evaluation Procedures for Data Streams, Testing Framework, Environments, Data Sources, Generation Speed and Data Size, Evolving Stream Experimental Setting.

UNIT- II
Hoeffding Trees, The Hoeffding Bound for Tree Induction, The Basic Algorithm, Memory Management, Numeric Attributes, Batch Setting Approaches, Data Stream Approaches.

UNIT- III
Prediction Strategies, Majority Class, Naïve Bayes Leaves, Adaptive Hybrid, Hoeffding Tree Ensembles, Data Stream Setting, Realistic Ensemble Sizes.

UNIT- IV
Evolving Data Streams, Algorithms for Mining with Change, A Methodology for Adaptive Stream Mining, Optimal Change Detector and Predictor, Adaptive Sliding Windows, Introduction, Maintaining Updated Windows of Varying Length.

UNIT- V
Adaptive Hoeffding Trees, Introduction, Decision Trees on Sliding Windows, Hoeffding Adaptive Trees, Adaptive Ensemble Methods, New methods of Bagging using trees of different size, New method of bagging using ADWIN, Adaptive Hoeffding Option Trees, Method performance.

TEXT BOOK:
1. DATA STREAM MINING: A Practical Approach by Albert Bifet and Richard Kirkby.

REFERENCE BOOKS:
1. Knowledge discovery from data streams by Gama João. ISBN: 978-1-4398-2611-9
2. Machine Learning for Data Streams by Albert Bifet, Ricard Gavalda; MIT Press, 2017