JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B.Tech in CSE (DATA SCIENCE)
III & IV YEAR COURSE STRUCTURE & TENTATIVE SYLLABUS (R18)

Applicable From 2020-21 Admitted Batch

III YEAR I SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1		Design and Analysis of Algorithms	3	0	0	3
2		Introduction to Data Science	3	0	0	3
3		Computer Networks	3	0	0	3
4		Data Mining	3	0	0	3
5		Professional Elective - I	3	0	0	3
6		Professional Elective - II	3	0	0	3
7		Data Mining Lab	0	0	3	1.5
8		Computer Networks Lab	0	0	3	1.5
9		Advanced Communication Skills Lab	0	0	2	1
10		Intellectual Property Rights	3	0	0	0
		Total Credits	**21**	**0**	**8**	**22**

III YEAR II SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1		Compiler Design	3	1	0	4
2		Machine Learning	3	1	0	4
3		Big Data Analytics	3	1	0	4
4		Professional Elective – III	3	0	0	3
5		Open Elective - I	3	0	0	3
6		Machine Learning Lab	0	0	3	1.5
7		Big Data Analytics Lab	0	0	3	1.5
8		Professional Elective - III Lab	0	0	2	1
9		Environmental Science	3	0	0	0
		Total Credits	**18**	**3**	**8**	**22**

IV YEAR I SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1		Predictive Analytics	3	0	0	3
2		Web and Social Media Analytics	2	0	0	2
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		Web and Social Media Analytics Lab	0	0	2	1
7		Industrial Oriented Mini Project/ Summer Internship	0	0	0	2*
8		Seminar	0	0	2	1
9		Project Stage – I	0	0	6	3
		Total Credits	**14**	**0**	**10**	**21**

IV YEAR II SEMESTER

S. No.	Course Code	Course Title	L	T	P	Credits
1		Organizational Behaviour	3	0	0	3
2		Professional Elective -VI	3	0	0	3
3		Open Elective-III	3	0	0	3
4		Project Stage - II	0	0	14	7
		Total Credits	**9**	**0**	**14**	**16**

*Note: Industrial Oriented Mini Project/ Summer Internship is to be carried out during the summer vacation between 6th and 7th semesters. Students should submit report of Industrial Oriented Mini Project/ Summer Internship for evaluation.

MC - Environmental Science – Should be Registered by Lateral Entry Students Only.
MC – Satisfactory/Unsatisfactory

Professional Elective-I

	Data Warehousing and Business Intelligence
	Artificial Intelligence
	Web Programming
	Image Processing
	Computer Graphics

Professional Elective - II

	Spatial and Multimedia Databases
	Information Retrieval Systems
	Software Project Management
	DevOps
	Computer Vision and Robotics

Professional Elective - III

	Software Testing Methodologies
	Data Visualization Techniques
	Scripting Languages
	Mobile Application Development
	Cryptography and Network Security

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

Professional Elective –IV

	Quantum Computing
	Database Security
	Natural Language Processing
	Information Storage Management
	Internet of Things

Professional Elective - V

	Privacy Preserving in Data Mining
	Cloud Computing
	Data Science Applications
	Mining Massive Datasets
	Exploratory Data Analysis

Professional Elective – VI

	Data Stream Mining
	Web Security
	Video Analytics
	Blockchain Technology
	Parallel and Distributed Computing

PREDICTIVE ANALYTICS

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

Course Objectives: The course serves to advance and refine expertise on theories, approaches and techniques related to prediction and forecasting.

Course Outcomes
1. Understand prediction-related principles, theories and approaches.
2. Learn model assessment and validation.
3. Understand the basics of predictive techniques and statistical approaches.
4. Analyze supervised and unsupervised algorithms.

UNIT - I
Linear Methods for Regression and Classification: Overview of supervised learning, Linear regression models and least squares, Multiple regression, Multiple outputs, Subset selection, Ridge regression, Lasso regression, Linear Discriminant Analysis, Logistic regression, Perceptron learning algorithm.

UNIT - II
Model Assessment and Selection: Bias, Variance, and model complexity, Bias-variance trade off, Optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters, Bayesian approach and BIC, Cross- validation, Boot strap methods, conditional or expected test error.

UNIT - III
Additive Models, Trees, and Boosting: Generalized additive models, Regression and classification trees, Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting, Examples (Spam data, California housing, New Zealand fish, Demographic data).

UNIT - IV
Neural Networks (NN), Support Vector Machines (SVM), and K-nearest Neighbor: Fitting neural networks, Back propagation, Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest – Neighbour classifiers (Image Scene Classification).

UNIT - V
Unsupervised Learning and Random forests: Association rules, Cluster analysis, Principal Components, Random forests and analysis.

TEXT BOOK:
1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning-Data Mining, Inference, and Prediction, Second Edition, Springer Verlag, 2009.

REFERENCE BOOKS:
1. C.M.Bishop –Pattern Recognition and Machine Learning, Springer, 2006.
2. L. Wasserman-All of statistics.
3. Gareth James. Daniela Witten. Trevor Hastie Robert Tibshirani. An Introduction to Statistical Learning with Applications in R.

WEB AND SOCIAL MEDIA ANALYTICS

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

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

Course Outcomes:
1. Knowledge on decision support systems.
2. Apply natural language processing concepts on text analytics.
3. Understand sentiment analysis.
4. 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.

QUANTUM COMPUTING (Professional Elective – IV)

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

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

Course Outcomes:
1. Understand basics of quantum computing
2. Understand physical implementation of Qubit
3. Understand Quantum algorithms and their implementation
4. Understand the Impact of Quantum Computing on Cryptography

UNIT - I
Introduction to Essential Linear Algebra: Some Basic Algebra, Matrix Math, Vectors and Vector Spaces, Set Theory. **Complex Numbers:** Definition of Complex Numbers, Algebra of Complex Numbers, Complex Numbers Graphically, Vector Representations of Complex Numbers, Pauli Matrice, Transcendental Numbers.

UNIT - II
Basic Physics for Quantum Computing: The Journey to Quantum, Quantum Physics Essentials, Basic Atomic Structure, Hilbert Spaces, Uncertainty, Quantum States, Entanglement.
Basic Quantum Theory: Further with Quantum Mechanics, Quantum Decoherence, Quantum Electrodynamics, Quantum Chromodynamics, Feynman Diagram Quantum Entanglement and QKD, Quantum Entanglement, Interpretation, QKE.

UNIT - III
Quantum Architecture: Further with Qubits, Quantum Gates, More with Gates, Quantum Circuits, The D-Wave Quantum Architecture. **Quantum Hardware:** Qubits, How Many Qubits Are Needed? Addressing Decoherence, Topological Quantum Computing, Quantum Essentials.

UNIT - IV
Quantum Algorithms: What Is an Algorithm? Deutsch's Algorithm, Deutsch-Jozsa Algorithm, Bernstein-Vazirani Algorithm, Simon's Algorithm, Shor's Algorithm, Grover's Algorithm.

UNIT - V
Current Asymmetric Algorithms: RSA, Diffie-Hellman, Elliptic Curve. **The Impact of Quantum Computing on Cryptography:** Asymmetric Cryptography, Specific Algorithms, Specific Applications.

TEXT BOOKS:
1. Nielsen M. A., Quantum Computation and Quantum Information, Cambridge University Press
2. Dr. Chuck Easttom, Quantum Computing Fundamentals, Pearson

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. Basic Concepts. Vol. Basic Tools and Special Topics, World Scientific.
3. Pittenger A. O., An Introduction to Quantum Computing Algorithms.

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 secure software design

Course Outcomes:
- Ability to carry out a risk analysis for large database.
- Ability to set up, and maintain the accounts with privileges and roles.

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 Hartson and Hsiao's Model Fernandez's Model Bussolati and Martella's Model for Distributed databases.

UNIT - II
Security Models -2: 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 conclusion.
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.

UNIT - III
Security Software Design: Introduction A Methodological Approach to Security Software Design, Secure Operating System Design, Secure DBMS Design Security Packages Database Security Design
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 - IV
Models for the Protection of New Generation Database Systems -1: 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.

UNIT - V
Models for the Protection of New Generation Database Systems -2: A Model for the Protection of New Generation Database Systems: the Orion Model ajodia and Kogan's Model A Model for the Protection of Active Databases Conclusions.

TEXT BOOKS:
1. Database Security by Castano, Pearson Edition
2. Database Security and Auditing: Protecting Data Integrity and Accessibility, 1st Edition, Hassan Afyouni, THOMSON Edition.

REFERENCE BOOK:
1. Database security by Alfred basta, melissazgola, CENGAGE learning.

NATURAL LANGUAGE PROCESSING (Professional Elective – IV)

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

Prerequisites: Data structures, finite automata and probability theory

Course Objectives:
- Introduce 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
- Able to 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

UNIT - II
Syntax Analysis: Parsing Natural Language, Treebanks: A Data-Driven Approach to Syntax, Representation of Syntactic Structure, Parsing Algorithms, Models for Ambiguity Resolution in Parsing, Multilingual Issues

UNIT - III
Semantic Parsing: Introduction, Semantic Interpretation, System Paradigms, Word Sense Systems, Software.

UNIT - IV
Predicate-Argument Structure, Meaning Representation Systems, Software.

UNIT - V
Discourse Processing: Cohension, Reference Resolution, Discourse Cohension and Structure
Language Modeling: Introduction, N-Gram Models, Language Model Evaluation, Parameter Estimation, Language Model Adaptation, Types of Language Models, Language-Specific Modeling Problems, 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.
2. Natural Language Processing and Information Retrieval: Tanvier Siddiqui, U.S. Tiwary.

REFERENCE BOOK:
1. Speech and Natural Language Processing - Daniel Jurafsky & James H Martin, Pearson Publications.

INFORMATION STORAGE MANAGEMENT (Professional Elective – IV)

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

Course Objectives:
1. To understand the basic components of Storage System Environment.
2. To understand the Storage Area Network Characteristics and Components.
3. To examine emerging technologies including IP-SAN.
4. To describe the different backup and recovery topologies and their role in providing disaster recovery and business continuity capabilities.
5. To understand the local and remote replication technologies.

Course Outcomes:
1. Understand the logical and physical components of a Storage infrastructure.
2. Evaluate storage architectures, including storage subsystems, DAS, SAN, NAS, and CAS.
3. Understand the various forms and types of Storage Virtualization.
4. Describe the different roles in providing disaster recovery and business continuity capabilities.
5. Distinguish different remote replication technologies.

UNIT - I
Introduction to Storage Technology: Data proliferation and the varying value of data with time & usage, Sources of data and states of data creation, Data center requirements and evolution to accommodate storage needs, Overview of basic storage management skills and activities, The five pillars of technology, Overview of storage infrastructure components, Evolution of storage, Information Lifecycle Management concept, Data categorization within an enterprise, Storage and Regulations.

UNIT - II
Storage Systems Architecture: Intelligent disk subsystems overview, Contrast of integrated vs. Modular arrays, Component architecture of intelligent disk subsystems, Disk physical structure-components, properties, performance, and specifications, Logical partitioning of disks, RAID & parity algorithms, hot sparing, Physical vs. logical disk organization, protection, and back end management, Array caching properties and algorithms, Front end connectivity and queuing properties, Front end to host storage provisioning, mapping, and operation, Interaction of file systems with storage, Storage system connectivity protocols.

UNIT - III
Introduction to Networked Storage: JBOD, DAS, SAN, NAS, & CAS evolution, Direct Attached Storage (DAS) environments: elements, connectivity, & management, Storage Area Networks (SAN): elements & connectivity, Fibre Channel principles, standards, & network management principles, SAN management principles, Network Attached Storage (NAS): elements, connectivity options, connectivity protocols (NFS, CIFS, ftp), & management principles, IP SAN elements, standards (iSCSI, FCIP, iFCP), connectivity principles, security, and management principles, Content Addressable Storage (CAS): elements, connectivity options, standards, and management principles, Hybrid Storage - solutions overview including technologies like virtualization & appliances.

UNIT - IV
Introductions to Information Availability: Business Continuity and Disaster Recovery Basics, Local business continuity techniques, Remote business continuity techniques, Disaster Recovery principles & techniques. **Managing & Monitoring:** Management philosophies (holistic vs. system & component), Industry management standards (SNMP, SMI-S, CIM), Standard framework applications, Key management metrics (thresholds, availability, capacity, security, performance), Metric analysis methodologies & trend analysis, Reactive and proactive management best practices, Provisioning &

configuration change planning, Problem reporting, prioritization, and handling techniques, Management tools overview.

UNIT - V
Securing Storage and Storage Virtualization: Define storage security. List the critical security attributes for information systems, describe the elements of a shared storage model and security extensions, Define storage security domains, List and analyze the common threats in each domain, Identify different virtualization technologies, describe block-level and file level virtualization technologies and processes.

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.

INTERNET OF THINGS (Professional Elective – IV)

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

Course Objectives:
1. To introduce the terminology, technology and its applications
2. To introduce the concept of M2M (machine to machine) with necessary protocols
3. To introduce the Python Scripting Language which is used in many IoT devices
4. To introduce the Raspberry PI platform, that is widely used in IoT applications
5. To introduce the implementation of web-based services on IoT devices

Course Outcomes:
1. Interpret the impact and challenges posed by IoT networks leading to new architectural models.
2. Compare and contrast the deployment of smart objects and the technologies to connect them to the network.
3. Appraise the role of IoT protocols for efficient network communication.
4. Elaborate the need for Data Analytics and Security in IoT.
5. Illustrate different sensor technologies for sensing real world entities and identify the applications of IoT in Industry.

UNIT - I

Introduction to Internet of Things –Definition and Characteristics of IoT, Physical Design of IoT – IoT Protocols, IoT communication models, Iot Communication APIs IoT enabaled Technologies – Wireless Sensor Networks, Cloud Computing, Big data analytics, Communication protocols, Embedded Systems, IoT Levels and Templates Domain Specific IoTs – Home, City, Environment, Energy, Retail, Logistics, Agriculture, Industry, health and Lifestyle

UNIT - II

IoT and M2M – Software defined networks, network function virtualization, difference between SDN and NFV for IoT Basics of IoT System Management with NETCOZF, YANG- NETCONF, YANG, SNMP NETOPEER

UNIT - III

Introduction to Python - Language features of Python, Data types, data structures, Control of flow, functions, modules, packaging, file handling, data/time operations, classes, Exception handling Python packages - JSON, XML, HTTPLib, URLLib, SMTPLib

UNIT - IV

IoT Physical Devices and Endpoints - Introduction to Raspberry PI-Interfaces (serial, SPI, I2C) Programming – Python program with Raspberry PI with focus of interfacing external gadgets, controlling output, reading input from pins.

UNIT - V

IoT Physical Servers and Cloud Offerings – Introduction to Cloud Storage models and communication APIs Webserver – Web server for IoT, Cloud for IoT, Python web application framework Designing a RESTful web API

TEXT BOOKS:
1. Internet of Things - A Hands-on Approach, Arshdeep Bahga and Vijay Madisetti, Universities Press, 2015, ISBN: 9788173719547.
2. Getting Started with Raspberry Pi, Matt Richardson & Shawn Wallace, O'Reilly (SPD), 2014, ISBN: 9789350239759.

PRIVACY PRESERVING IN DATA MINING (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:
1. The aim of the course is to introduce the fundamentals of Privacy Preserving Data Mining Methods
2. 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:
1. Understand the concepts of Privacy Preserving Data Mining Models and Algorithms.
2. Demonstrate a comprehensive understanding of different tasks associated in Inference Control Methods for Privacy-Preserving Data Mining.
3. Understand the concepts of Data Anonymization Methods and its Measures.
4. Evaluate and Appraise the solution designed for Multiplicative Perturbation.
5. 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. J. 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.

CLOUD COMPUTING (Professional Elective – V)

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

Pre-requisites: Courses on Computer Networks, Operating Systems, Distributed Systems.

Course Objectives:
1. This course provides an insight into cloud computing.
2. Topics covered include- distributed system models, different cloud service models, service-oriented architectures, cloud programming and software environments, resource management.

Course Outcomes:
1. Ability to understand various service delivery models of a cloud computing architecture.
2. Ability to understand the ways in which the cloud can be programmed and deployed.
3. Understanding cloud service providers.

UNIT - I
Computing Paradigms: High-Performance Computing, Parallel Computing, Distributed Computing, Cluster Computing, Grid Computing, Cloud Computing, Bio computing, Mobile Computing, Quantum Computing, Optical Computing, Nano computing.

UNIT - II
Cloud Computing Fundamentals: Motivation for Cloud Computing, The Need   for Cloud Computing, Defining Cloud Computing, Definition of Cloud computing, Cloud Computing Is a Service, Cloud Computing Is a Platform, Principles of Cloud computing, Five Essential Characteristics, Four Cloud Deployment Models.

UNIT - III
Cloud Computing Architecture and Management: Cloud architecture, Layer, Anatomy of the Cloud, Network Connectivity in Cloud Computing, Applications, on the Cloud, Managing the Cloud, Managing the Cloud Infrastructure Managing the Cloud application, Migrating Application to Cloud, Phases of Cloud Migration Approaches for Cloud Migration.

UNIT - IV
Cloud Service Models: Infrastructure as a Service, Characteristics of IaaS. Suitability of IaaS, Pros and Cons of IaaS, Summary of IaaS Providers, Platform as a Service, Characteristics of PaaS, Suitability of PaaS, Pros and Cons of PaaS, Summary of PaaS Providers, Software as a Service, Characteristics of SaaS, Suitability of SaaS, Pros and Cons of SaaS, Summary of SaaS Providers, Other Cloud Service Models.

UNIT V
Cloud Service Providers: EMC, EMC IT, Captiva Cloud Toolkit, Google, Cloud Platform, Cloud Storage, Google Cloud Connect, Google Cloud Print, Google App Engine, Amazon Web Services, Amazon Elastic Compute Cloud, Amazon Simple Storage Service, Amazon Simple Queue ,service, Microsoft, Windows Azure, Microsoft Assessment and Planning Toolkit, SharePoint, IBM, Cloud Models, IBM Smart Cloud, SAP Labs, SAP HANA Cloud Platform, Virtualization Services Provided by SAP, Sales force, Sales Cloud, Service Cloud: Knowledge as a Service, Rack space, VMware, Manjra soft, Aneka Platform.

TEXT BOOK:
1. Essentials of cloud Computing: K. Chandrasekhran, CRC press, 2014
REFERENCE BOOKS:
1. Cloud Computing: Principles and Paradigms by Rajkumar Buyya, James Broberg and Andrzej M. Goscinski, Wiley, 2011.
2. Distributed and Cloud Computing, Kai Hwang, Geoffery C. Fox, Jack J. Dongarra, Elsevier, 2012.
3. Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, Tim Mather, Subra Kumaraswamy, Shahed Latif, O'Reilly, SPD, rp 2011.

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: After completion of course, students would:
1. To correlate data science and solutions to modern problems.
2. To 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.

MINING MASSIVE DATASETS (Professional Elective – V)

B.Tech. IV Year I Sem.

	L	T	P	C
	3	0	0	3

Prerequisites: 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.

EXPLORATORY DATA ANALYSIS (Professional Elective – V)

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

Course Objectives:
1. This course introduces the methods for data preparation and data understanding.
2. It covers essential exploratory techniques for understanding multivariate data by summarizing it through statistical methods and graphical methods.
3. Supports to Summarize the insurers use of predictive analytics, data science and Data Visualization.

Course Outcomes:
1. Handle missing data in the real-world data sets by choosing appropriate methods.
2. Summarize the data using basic statistics. Visualize the data using basic graphs and plots.
3. Identify the outliers if any in the data set.
4. Choose appropriate feature selection and dimensionality reduction.
5. Techniques for handling multi-dimensional data.

UNIT - I:
Introduction to Exploratory Data Analysis: Data Analytics lifecycle, Exploratory Data Analysis (EDA)– Definition, Motivation, Steps in data exploration, The basic data types Data Type Portability.

UNIT - II:
Preprocessing - Traditional Methods and Maximum Likelihood Estimation: Introduction to Missing data, Traditional methods for dealing with missing data, Maximum Likelihood Estimation – Basics, Missing data handling, Improving the accuracy of analysis. **Preprocessing Bayesian Estimation:** Introduction to Bayesian Estimation, Multiple Imputation-Imputation Phase, Analysis and Pooling Phase, Practical Issues in Multiple Imputation, Models for Missing Notation Random Data.

UNIT - III:
Data Summarization & Visualization: Statistical data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, N-D Statistical data analysis.

UNIT - IV:
Outlier Analysis: Introduction, Extreme Value Analysis, Clustering based, Distance Based and Density Based outlier analysis, Outlier Detection in Categorical Data. **Feature Subset Selection:** Feature selection algorithms: filter methods, wrapper methods and embedded methods, Forward selection backward elimination, Relief, greedy selection, genetic algorithms for features selection.

UNIT - V
Dimensionality Reduction: Introduction, Principal Component Analysis (PCA), Kernel PCA, Canonical Correlation Analysis, Factor Analysis, Multidimensional scaling, Correspondence Analysis.

TEXT BOOKS:
1. Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining, by Glenn J. Myatt.

REFERENCE BOOKS:
1. Charu C. Aggarwal, "Data Mining The Text book", Springer, 2015.
2. Craig K. Enders, "Applied Missing Data Analysis", The Guilford Press, 2010.
3. Inge Koch, "Analysis of Multivariate and High dimensional data", Cambridge University Press, 2014.
4. Michael Jambu, "Exploratory and multivariate data analysis", Academic Press Inc., 1990.
5. Charu C. Aggarwal, "Data Classification Algorithms and Applications", CRC press, 2015.

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:
1. Knowledge on decision support systems.
2. Apply natural language processing concepts on text analytics.
3. Understand sentiment analysis.
4. 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 BOOKS:
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.