JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

R25 M.TECH DATA SCIENCE COURSE STRUCTURE AND SYLLABUS

Vision and Mission of the University

VISION

The University is primarily promoting quality of education in the areas of Science, Technology, Engineering and Mathematics (STEM) as four academic pillars of education, to excel in teaching, learning, research, consultancy and placements through innovative practices with global perspective.

MISSION

1. Design an Industry relevant curriculum from time to time with a Global perspective
2. Promoting quality education by embracing ICT delivery mechanism with continuous pedagogy through e-learning mechanism
3. Spread across for industry collaborations with a focus to pre-training and placements for technology transfer to society
4. Establishing centers of excellence to promote research and innovations in multidisciplinary areas to bring in patent culture and consultancy practices
5. International Collaborations for student outreach
6. Facilitating international students to study in JNTUK to infuse cross culture learning practices.

Vision and Mission of the Institute

Vision and Mission of the Department

Programme Education Objectives (PEOs) of the M. Tech (DS)

PEO 1: Build a strong foundation in data science concepts, including data analytics, machine learning, and statistical modeling, to develop intelligent systems that solve real-world problems across diverse domains.

PEO 2: Promote continuous learning, innovation, and research by exploring emerging technologies in data science, and contribute to academic, industrial, and societal advancements through data-driven solutions.

PEO 3: Apply ethical practices, communication, teamwork, and managerial skills to function effectively in multidisciplinary environments as data scientists, researchers, consultants, or entrepreneurs.

Mapping of Mission statements to PEOs

Programme Outcomes (POs)

PO1: An ability to independently carry out research /investigation and development work to solve practical problems

PO2: An ability to write and present a substantial technical report/document

PO3: Students should be able to demonstrate a degree of mastery over the area as per the specialization of the program. The mastery should be at a level higher than the requirements in the appropriate bachelor program

PO4: Demonstrate a strong foundation in mathematics, statistics, machine learning, and core data science concepts to analyse and interpret complex datasets.

PO5: Design and implement scalable data-driven solutions to real-world problems using advanced tools and techniques in data analytics, artificial intelligence, and big data technologies.

PO6: Apply appropriate research methodologies to explore challengeable problems in data science, and develop innovative solutions using scientific thinking and experimentation.

Note: Program may add up to three additional POs

Mapping of Programme Outcomes to PEOs

Mapping of Programme Outcomes to GAs

M. Tech

DATA SCIENCE

Programme Course Structure & Syllabus

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

R25 M.TECH DATA SCIENCE COURSE STRUCTURE AND SYLLABUS

Programme Structure R25 M. Tech (DS) Structure

I Semester

S.No.	Course Title	L	T	P	C
1	**Program Core – 1** Advanced Algorithms Analysis	3	1	0	4
2	**Program Core – 2** Data Science and its Applications	3	1	0	4
3	**Program Core – 3** Artificial Intelligence	3	1	0	4
4	**Program Elective – I**	3	0	0	3
5	**Program Elective – II**	3	0	0	3
6	**Laboratory – 1** Advanced Algorithms analysis lab	0	1	2	2
7	**Laboratory – 2** Data Science Applications using R/Python Lab	0	1	2	2
8	Seminar-I	0	0	2	1
	TOTAL	**15**	**5**	**6**	**23**

List of Professional Elective Courses in I Semester (Electives – I & II)

S.No.	Course Title
1	Massive Mining Datasets
2	Full Stack Technologies
3	Agile Methodologies
4	Feature Engineering
5	Cryptography and network Security
6	Information Retrieval Systems
7	Natural Language Processing
8	Data Engineering

@ Minimum 2/3 themes per elective

II Semester

S. No.	Course Title	L	T	P	C
1	**Program Core – 4** Large Language Models	3	1	0	4
2	**Program Core – 5** Machine Learning	3	1	0	4
3	**Program Core – 6** Foundations to Data Analytics	3	1	0	4
4	**Program Elective – III**	3	0	0	3
5	**Program Elective – IV**	3	0	0	3
6	**Laboratory – 3** Large Language Models Lab	0	1	2	2
7	**Laboratory – 4** Machine Learning Lab	0	1	2	2
8	Seminar – II	0	0	2	1
	TOTAL	**15**	**5**	**6**	**23**

List of Professional Elective Courses in II Semester (Electives III & IV)

S.No.	Course Title
1	Deep Learning
2	Database and Web Application Security
3	Software Project Management
4	Big Data Analytics
5	Object Oriented Analysis & Design
6	Quantum Computing
7	Information Theory & coding
8	Block Chain Technologies

@ Minimum 2/3 themes per elective

III Semester

S. No.	Course Title	L	T	P	C
1	Research Methodology and IPR / *Swayam 12 week MOOC course – RM&IPR*	3	0	0	3
2	Summer Internship/ Industrial Training (8-10 weeks)*	-	-	-	3
3	Comprehensive Viva#	-	-	-	2
4	Dissertation Part – A$	-	-	20	10
	TOTAL	**3**	**-**	**20**	**18**

*** Student attended during summer / year break and assessment will be done in 3rd Sem. # Comprehensive viva can be conducted courses completed upto second sem. $ Dissertation – Part A, internal assessment**

IV Semester

S. No.	Course Title	L	T	P	C
1	Dissertation Part – B%	-	-	32	16
	TOTAL	**-**	**-**	**32**	**16**

% External Assessment

I semester	ADVANCED ALGORITHMS ANALYSIS	L	T	P	C
		3	1	0	4

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Analyze the complexity/performance of different algorithms.	K4
CO2	Determine the appropriate data structure for solving a particular set of problems.	K3
CO3	Categorize the different problems in various classes according to their complexity	K4
CO4	Discuss Dynamic Programming and Fast Fourier Transform algorithm	K4
CO5	Explain Linear Programming	K5

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	2	1
CO2	2		3	3	2	1
CO3	2	1	3		2	1
CO4	2	1		2	2	
CO5	2	1	3	3	2	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	Contact Hours
UNIT – 1	**Sorting:** Review of various sorting algorithms, topological sorting Graph: Definitions and Elementary Algorithms: Shortest path by BFS, shortest path in edge-weighted case (Dijkstra's), depth-first search and computation of strongly connected components, emphasis on correctness proof of the algorithm and time/space analysis, example of amortized analysis.	10Hrs
UNIT – 2	**Matroids**: Introduction to greedy paradigm, algorithm to compute a maximum weight maximal independent set. Application to MST. Graph Matching: Algorithm to compute maximum matching. Characterization of maximum matching by augmenting paths, Edmond's Blossom algorithm to compute augmenting path.	12Hrs

UNIT – 3	**Flow-Networks:** Maxflow-mincut theorem, Ford-Fulkerson Method to compute maximum flow, Edmond-Karp maximum-flow algorithm. Matrix Computations: Strassen's algorithm and introduction to divide and conquer paradigm, inverse of a triangular matrix, relation between the time complexities of basic matrix operations, LUPdecomposition.	12Hrs
UNIT – 4	**Shortest Path in Graphs:** Floyd-Warshall algorithm and introduction to dynamic programming paradigm. More examples of dynamic programming. Modulo Representation of integers/polynomials: Chinese Remainder Theorem, Conversion between base-representation and modulo-representation. Extension to polynomials. Application: Interpolation problem. Discrete Fourier Transform (DFT): In complex field, DFT in modulo ring. Fast Fourier Transform algorithm. Schonhage-Strassen Integer Multiplication algorithm	12Hrs
UNIT – 5	**Linear Programming:** Geometry of the feasibility region and Simplex algorithm NP-completeness: Examples, proof of NP-hardness and NP-completeness. Recent Trends in problem solving paradigms using recent searching and sorting techniques by applying recently proposed data structures.	12Hrs
	Total	58Hrs

*Note:

REFERENCES:

1. Cormen, Leiserson, Rivest, Stein, "Introduction to Algorithms".
2. Aho, Hopcroft, Ullman "The Design and Analysis of Computer Algorithms".
3. Kleinberg and Tardos."Algorithm Design".

I Semester	DATA SCIENCE AND ITS APPLICATIONS	L	T	P	C
		3	1	0	4

Course Objectives:

- Provide you with the knowledge and expertise to become a proficient data scientist.
- Demonstrate an understanding of statistics and machine learning concepts that are vital for data science.
- Produce Python code to statistically analyze a dataset.
- Critically evaluate data visualizations based on their design and use for communicating stories from data**.**

Course Outcomes:

		Knowledge Level (K)#
CO1	Explain how data is collected, managed and stored for data science.	K3
CO2	Understand the key concepts in data science, including their real-world applications and the toolkit used by data scientists.	K2
CO3	Discuss different Packages and how to use it	K5
CO4	Implement data collection and management scripts using Python Pandas.	K6

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	3	2
CO2	2	1	3	3	3	2
CO3	1	1	2	2	3	1
CO4	2	1	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**PYTHON Basics and Programming Concepts:** Introducing Python, Types and Operations -Numbers, Strings, Lists, Tuples, Dictionaries, Files, Numeric Types, Dynamic Typing; Statements and Syntax - Assignments, Expressions, Statements, Loops, iterations, comprehensions; Functions – Function Basics, Scopes, Arguments, Advanced Functions; Modules - Module Coding Basics, Module Packages, Advanced Module Topics; Classes and OOP - Class, Operator Overloading, Class Designing; Exceptions and Tools - Exception Basics,	10Hrs

	Exception Coding Details, Exception Objects, Designing With Exceptions, Parallel System Tools.	
UNIT – 2	**GUI Programming: Graphical User Interface -** Python GUI development options, Adding Widgets, GUI Coding Techniques, Customizing Widgets; Internet Programming - Network Scripting, Client-Side scripting, Pymailgui client, server-side scripting, Pymailcgi server; Tools and Techniques - databases and persistence, data structures, text and language, python/c integration**.**	10Hrs
UNIT – 3	**Pandas and NumPy: Numpy Basics -** Fast Element wise array functions, Multidimensional Array, Data Processing using arrays, file i/o with arrays; Pandas - Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics, Handling Missing Data, Hierarchical Indexing.	12Hrs
UNIT – 4	**Data Preprocessing:** Data Loading, Storage, and FileFormats - Reading and Writing data in text format, binary data formats, interacting with html and web apis, interacting with databases; Data Wrangling: Clean, Transform, Merge, Reshape - Combining and Merging Data Sets, Reshaping and Pivoting, Data Transformation, String Manipulation; Data Aggregation and Group Operations – Group by Mechanics, Data Aggregation, Groupby Operations and and Transformations, Pivot Tables and Cross Tabulation.	12Hrs
UNIT – 5	**Data Visualization:** A Brief matplotlib API Primer, Plotting Functions in pandas, Time Series, Financial and Economic Data Applications	12Hrs
	Total	56Hrs

*Note:

Text Books:

1. Learning Python , OReilly, Mark Lutz
2. Programming Python, OReilly, Mark Lutz
3. Python For Data Analysis (O Reilly, Wes Mckinney)

Reference Books:

1. Python: The Complete Reference, Martin C. Brown, McGraw Hill Education
2. Head First Python, Paul Barry, O'Reilly

I Semester	ARTIFICIAL INTELLIGENCE	L	T	P	C
		3	1	0	4

Course Outcomes:

		Knowledge Level (K)#
CO1	Demonstrate knowledge of the building blocks of AI as presented in terms of intelligent agents	K3
CO2	Analyze and formalize the problem as a state space, graph, design heuristics and select amongst different search or game based techniques to solve them	K3
CO3	Develop intelligent algorithms for constraint satisfaction problems and also design intelligent systems for Game Playing.	K6
CO4	Attain the capability to represent various real life problem domains using logic based techniques and use this to perform inference or planning.	K2
CO5	Solve problems with uncertain information using Bayesian approaches	K2

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	2	1
CO2	3	1	3	3	3	2
CO3	3	1	3	3	3	2
CO4	3		3	3	3	2
CO5	3	1	3	3	3	3

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Introduction to artificial intelligence**: Introduction, history, intelligent systems, foundations of AI, applications, tic-tac-tie game playing, development of AI languages, current trends in AI, Problem solving: state-space search and control strategies: Introduction, general problem solving, characteristics of problem, exhaustive searches, heuristic search techniques, iterative-deepening A*, constraint satisfaction.	10Hrs
UNIT – 2	**Problem reduction and game playing:** Introduction, problem reduction, game playing, alpha-beta pruning, two-player perfect information games, Logic concepts: Introduction, propositional calculus, proportional logic, natural deduction system, axiomatic	10Hrs

	system, semantic tableau system in proportional logic, resolution refutation in proportional logic, predicate logic	
UNIT – 3	**Knowledge representation**: Introduction, approaches to knowledge representation, knowledge representation using semantic network, extended semantic networks for KR, knowledge representation using frames, advanced knowledge representation techniques: Introduction, conceptual dependency theory, script structure, cyc theory, case grammars, semantic web.	12Hrs
UNIT – 4	**Uncertainty measure**: probability theory: Introduction, probability theory, Bayesian belief networks (exact and approximate inference) certainty factor theory, dempster-shafer theory, non-monotonic reasoning, TMS.	12Hrs
UNIT – 5	**Expert systems:-** Introduction, basic concepts, structure of expert systems, the human element in expert systems how expert systems works, problem areas addressed by expert systems, expert systems success factors, types of expert systems, expert systems and the internet interacts web, knowledge engineering, scope of knowledge, difficulties, in knowledge acquisition methods of knowledge acquisition, machine learning, intelligent agents, selecting an appropriate knowledge acquisition method, societal impacts reasoning in artificial intelligence, inference with rules, with frames: model based reasoning, case based reasoning, explanation & meta knowledge inference with uncertainty representing uncertainty.	12Hrs
	Total	56Hrs

Text Books:

1. Artificial intelligence, A modern Approach, 2nded, Stuart Russel, Peter Norvig, Prentice Hall
2. Artificial Intelligence, SarojKaushik, 1st Edition, CENGAGE Learning, 2011.

Reference Books:

1. Artificial intelligence, structures and Strategies for Complex problem solving, 5th
2. Edition, George F Lugar, PEA
3. Introduction to Artificial Intelligence, Ertel, Wolf Gang, Springer, 2017
4. Artificial Intelligence, A new Synthesis, 1st Edition, Nils J Nilsson, Elsevier,1998
5. Artificial Intelligence- 3rd Edition, Rich, Kevin Knight, Shiv Shankar B Nair, TMH
6. Introduction To Artificial Intelligence And Expert Systems, 1st Edition,Patterson, Pearson India, 2015

I Semester	MASSIVE MINING DATASETS	L	T	P	C
		3	0	0	3

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Recollecting fundamentals of data mining- L2	K2
CO2	Apply the concept of Map reduce and data streams for storing and processing of massive data sets - L3	K3
CO3	Analyze the issues underlying the effective applications of massive data sets -L4	K4
CO4	Evaluate different clustering algorithms and analyze various decomposition techniques - L4	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	1	1	2	2	2	1
CO2	2	1	3	3	3	2
CO3	3	1	3	3	3	2
CO4	3	1	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Data Mining:** Introduction, Statistical Modeling, Machine Learning, Computational Approaches to Modeling, Feature Extraction, Statistical Limits on Data Mining, Hash Functions, Indexes, Natural Logarithms, Power Laws. (CO 1)	10Hrs
UNIT – 2	**Map Reduce and the New Software Stack:** Distributed File Systems, Map Reduce, Algorithms Using Map Reduce, Extensions to Map Reduce, Complexity Theory for Map Reduce (CO 2)	12Hrs
UNIT – 3	**Mining Data Streams:** The Stream Data Model, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream, Counting Ones in a Window, Decaying Windows. (CO 1, CO 2)	12Hrs
UNIT – 4	**Frequent Item sets:** The Market-Basket Model, Market Baskets and the A-Priori Algorithm, Handling Larger Datasets in Main Memory, Limited-Pass Algorithms, Counting Frequent Items in a Stream. (CO 1, CO 3)	12Hrs

UNIT – 5	**Clustering:** Introduction to Clustering Techniques, Hierarchical Clustering, K-means Algorithms, The CURE Algorithm, Clustering in Non-Euclidean Spaces, and Clustering for Streams and Parallelism. Dimensionality Reduction: Eigen values and Eigenvectors of Symmetric Matrices, Principal-Component Analysis, Singular-Value Decomposition, CUR Decomposition (CO 1, CO 4).	12Hrs
	Total	58Hrs

*Note:

Text Books:

1. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, 2nd Edition, Daniel Jurafsky, James H. Martin - Pearson Publication,2014.
2. Natural Language Processing with Python, First Edition, Steven Bird, Ewan Klein and Edward Loper, OReilly Media,2009.

Reference Books:

1. Language Processing with Java and Ling Pipe Cookbook, 1st Edition, Breck Baldwin, Atlantic Publisher, 2015.
2. Natural Language Processing with Java, 2nd Edition, Richard M Reese, OReilly Media,2015.
3. Handbook of Natural Language Processing, Second, Nitin Indurkhya and Fred J. Damerau, Chapman and Hall/CRC Press, 2010.Edition

   Natural Language Processing and Information Retrieval, 3rd Edition, Tanveer Siddiqui, U.S. Tiwary, Oxford University Press,2008.

I Semester	FULL STACK TECHNOLOGIES	L	T	P	C
		3	0	0	3

Course Objectives:

Translate user requirements into the overall architecture and implementation of new systems and Manage Project and coordinate with the Client.

- Write backend code in Python/Java, PHP languages and Writing optimized front end code HTML and JavaScript.
- Understand, create and debug database related queries and Create test code to validate the applications against client requirement.
- Monitor the performance of web applications & infrastructure and Troubleshooting web application with a fast and accurate a resolution

Course Outcomes: At the end of the

		Knowledge Level (K)#
CO1	Understand the architecture of the web,HTML5,CSS3,XML,DOM and related web standards and protoclos	K2
CO2	Apply JavaScript and AngularJS to create dynamic and responsive web applications	K3
CO3	Develop backend web applications using PHP and Node.js with form handling and database connectivity	K3
CO4	Perform database operations using MySQL and integrate it with frontend technologies using jQuery	K4
CO5	Demonstrate the use of MongoDB, deploy web applications, and implement web services using SOAP and RESTful APIs.	K5

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	2	2	1
CO2	2	1	3	3	2	2
CO3	3	1	3	3	3	2
CO4	2	1	3	3	3	2
CO5	2	2	3	3	3	3

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Introduction to web-** Internet and world wide web, Domain name service, Protocols: HTTP, FTP, SMTP, Html5 concepts, CSS3, Anatomy of a web page, XML- Document type Definition, XML schemas, Document object model, XSLT, DOM and SAX Approaches.	10Hrs
UNIT – 2	**Javascript-** The Basics of Javascript: Objects, Primitives Operations and Expressions, Control Statements, Arrays, Functions, Constructors, Patttern Matching using Regular Expressions, Angular Java Script- AngularJS Expressions: ARRAY, Objects, $eval, Strings, AngularJS Form Validation & Form Submission.	10Hrs
UNIT – 3	**PHP Programming: Back** – end- Scripts PHP, Node js, Working with PHP- Using variables, Using constants, Data types, Operators. Conditional & Control statements, Arrays, functions. Working with forms and Databases such as MySQL, Node.js- Introduction, Advantages, Node.js Process Model, Node JS Modules.	12Hrs
UNIT – 4	**JQuery:** Introduction to JQuery, Syntax, Selectors & Events, MySQL: Practice MySQL Queries, Aggregate Functions, Regular Expressions, Joins & Unions, Sub-Queries, Database Connectivity with MySql.	12Hrs
UNIT – 5	**Mongo DB-** Introduction, Architecture, Features, Examples, Database Creation & Collection in Mongo DB, Deploying Applications- Web hosting & Domains, Deployment Using Cloud Platforms, Web Services- SOAP, WSDL and RESTful Architecture.	12Hrs
	Total	56Hrs

*Note:

Text Books:

1. Programming the World Wide Web, 7 th Edition, Robet W Sebesta, Pearson, 2013
2. Web Technologies, 1st Edition 7th impression, Uttam K Roy, Oxford, 2012.
3. Pro Mean Stack Development, 1 st Edition, ELad Elrom, Apress O'Reilly, 2016
4. JavaScript & jQuery the missing manual, 2 nd Edition, David sawyer mcfarland, O'Reilly, 2011
5. Web Hosting for Dummies, 1st Edition, Peter Pollock, John Wiley & Sons, 2013
6. RESTful web services, 1 st Edition, Leonard Richardson, Ruby, O'Reilly, 2007

Reference Books:

1. Ruby on Rails Up and Running, Lightning fast Web development, 1 st Edition, Bruce Tate, Curt Hibbs, Oreilly, 2006
2. Programming Perl, 4 th edition, Tom Christiansen, Jonathan Orwant, O'Reilly ,2012
3. Web Technologies, HTML< JavaScript, PHP, Java, JSP, XML and AJAX, Black book, 1st Edition, Dream Tech, 2009
4. An Introduction to Web Design, Programming, 1 st Edition, Paul S Wang, Sanda S Katila, Cengage Learning, 2003

Web Reference Links:1. http://www.upriss.org.uk/perl/PerlCourse.html

I Semester	AGILE METHODOLOGIES (PROGRAM ELECTIVE-II)	L	T	P	C
		3	0	0	3

Course Objectives: This course is aimed at enabling the students to

1. The main objectives of this course are to introduce the important concepts of Agile software development Process, emphasize the role of stand-up meetings in software collaboration, impart the knowledge on values and principles in understanding agility

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Understand the core values and mindset of Agile Methodology for effective project development	K1
CO2	Explain Agile Principles and apply them in Agile Project management practices	K3
CO3	Describe Key concepts of XP, Simplicity, and Incremental Design	K2
CO4	Apply Lean Principles to identify and Eliminating Waste in software processes	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	2	2	1
CO2	2	2	3	2	3	2
CO3	2	1	2	2	2	1
CO4	3	1	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Learning Agile**: Getting Agile into your brain, Understanding Agile values, No Silver Bullet, Agile to the Rescue, adding Agile makes a difference. A fractured perspective, How a fractured perspective causes project problems. The Agile Manifesto, Purpose behind Each Practice. Individuals and Interactions Over Processes and Tools, Working Software over Comprehensive Documentation, Customer Collaboration over Contract Negotiation, Responding to Change over Following a Plan, Principles over Practices. Understanding the Elephant, Methodologies Help You Get It All in Place at Once, Where to Start with a New Methodology	10Hrs

UNIT – 2	**The Agile Principles:** The 12 Principles of Agile Software, The Customer Is Always Right, "Do As I Say, Not As I Said". Delivering the Project, Better Project Delivery for the Ebook Reader Project. Communicating and Working Together, Better Communication for the Ebook Reader Project. Project Execution—Moving the Project Along, A Better Working Environment for the Ebook Reader Project Team. Constantly Improving the Project and the Team. The Agile Project: Bringing All the Principles Together	12Hrs
UNIT – 3	**SCRUM and Self-Organizing Teams:** The Rules of Scrum, Act I: I Can Haz Scrum?, Everyone on a Scrum Team owns the Project, The Scrum Master Guides the Team's Decisions, The Product Owner Helps the Team Understand the Value of the Software, Everyone Owns the Project, Scrum Has Its Own Set of Values ,Status Updates Are for Social Networks!, The Whole Team Uses the Daily Scrum, Feedback and the Visibility-Inspection-Adaptation Cycle, The Last Responsible Moment, How to Hold an Effective Daily Scrum. Sprinting into a Wall, Sprints, Planning, and Retrospectives, Iterative or Incremental?, The Product Owner Makes or Breaks the Sprint, Visibility and Value, How to Plan and Run an Effective Scrum Sprint **Scrum Planning And Collective Commitment:** Not Quite Expecting the Unexpected, User Stories, Velocity, and Generally Accepted Scrum Practices, Make Your Software Useful, User Stories Help Build Features Your Users Will Use, Conditions of Satisfaction, Story Points and Velocity, Burndown Charts, Planning and Running a Sprint Using Stories, Points, Tasks, and a Task Board. Victory Lap, Scrum Values Revisited, Practices Do Work Without the Values (Just Don't Call It Scrum), Is Your Company's Culture Compatible with Scrum Values.	12Hrs
UNIT – 4	**XP And Embracing Change:** Going into Overtime, The Primary Practices of XP, Programming Practices, Integration Practices, Planning Practices, Team Practices, Why Teams Resist Changes, and How the Practices Help. The Game Plan Changed, but We're Still Losing, The XP Values Help the Team Change Their Mindset, XP Helps Developers Learn to Work with Users, Practices Only "Stick" When the Team Truly Believes in Them, An Effective Mindset Starts with the XP Values, The XP Values, Paved with Good Intentions. The Momentum Shifts, Understanding the XP Principles Helps You Embrace Change, The Principles of XP, XP Principles Help You Understand Planning, XP Principles Help You Understand Practices—and Vice Versa, Feedback Loops. **XP, Simplicity, and Incremental Design:** Code and Design, Code Smells and Antipatterns (or, How to Tell If You're Being Too Clever), XP Teams Look for Code Smells and Fix Them, Hooks, Edge Cases,	12Hrs

	and Code That Does Too Much. Make Code and Design Decisions at the Last Responsible Moment, Fix Technical Debt by Refactoring Mercilessly, Use Continuous Integration to Find Design Problems, Avoid Monolithic Design, Incremental Design and the Holistic XP Practices. Teams Work Best When They Feel Like They Have Time to Think, Team Members Trust Each Other and Make Decisions Together. The XP Design, Planning, Team, and Holistic Practices Form an Ecosystem Incremental Design Versus Designing for Reuse, When Units Interact in a Simple Way, the System Can Grow Incrementally, Great Design Emerges from Simple Interactions, Final Score.	
UNIT – 5	**Lean, Eliminating Waste, and Seeing the whole:** Lean Thinking, Commitment, Options Thinking, and Set-Based Development, Creating Heroes and Magical Thinking. Eliminate Waste, Use a Value Stream Map to Help See Waste Clearly, Gain a Deeper Understanding of the Product, See the Whole, Find the Root Cause of Problems That You Discover. Deliver As Fast As Possible, Use an Area Chart to Visualize Work in Progress, Control Bottlenecks by Limiting Work in Progress. **Kanban, Flow, and Constantly Improving:** The Principles of Kanban, Find a Starting Point and Evolve Experimentally from There. Stories Go into the System; Code Comes Out, Improving Your Process with Kanban, Visualize the Workflow, Limit Work in Progress. Measure and Manage Flow, Managing Flow with WIP Limits Naturally Creates Slack. Make Process Policies Explicit So Everyone Is on the Same Page. Emergent Behavior with Kanban. **The Agile Coach:** Coaches Understand Why People Don't Always Want to Change. The Principles of Coaching.	12Hrs
	Total	58Hrs

*Note:

Text Books:

1. Andrew Stellman, Jill Alison Hart, Learning Agile, O'Reilly, 2015.

Reference Books:

1. Andrew stellman, Jennifer Green, Head first Agile, O'Reilly, 2017.
2. Rubin K , Essential Scrum : A practical guide to the most popular Agile process, Addison-Wesley, 2013

I Semester	FEATURE ENGINEERING (PROGRAM ELECTIVE-III)	L	T	P	C
		3	0	0	3

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Describe the Basic concepts of Data, Tasks, Models, Features and Model building	K2
CO2	Explain the concept of coverting Text into Flat Vectors using Bag- of-Words, and Bag-of-n-Grams	K3
CO3	Demonstrate techniques for Dimensionality Reduction	K4
CO4	Discuss non linear Featurization	K4
C05	Explain the concept of Item-Based Collaborative Filtering	K3

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	2	1
CO2	2	1	3	3	3	2
CO3	3	1	3	3	3	2
CO4	2	1	3	3	3	2
CO5	3	1	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**The Machine Learning Pipeline:** Data, Tasks, Models, Features, Model Evaluation Fancy Tricks with Simple Numbers: Scalars, Vectors, and Spaces, Dealing with Counts, Binarization, Quantization or Binning, Log Transformation, Log Transform in Action, Power Transforms: Generalization of the Log Transform, Feature Scaling or Normalization, Min-Max Scaling, Standardization (Variance Scaling), ℓ2 Normalization, Interaction Features, Feature Selection	10Hrs
UNIT – 2	**Text Data: Flattening, Filtering, and Chunking: Bag-of-X:** Turning Natural Text into Flat Vectors, Bag- of-Words, Bag-of-n-Grams, Filtering for Cleaner Features: Stopwords, Frequency-Based Filtering, Stemming; Atoms of Meaning: From Words to n-Grams to Phrases: Parsing and Tokenization, Collocation Extraction for Phrase Detection The Effects of Feature Scaling: From Bag-of-Words to Tf-Idf :Tf-Idf : A Simple Twist on Bag-of- Words, Putting It to the Test : Creating a	12Hrs

	Classification Dataset, Scaling Bag-of-Words with Tf-Idf Transformation, Classification with Logistic Regression, Tuning Logistic Regression with Regularization	
UNIT – 3	**Categorical Variables:** Counting Eggs in the Age of Robotic Chickens: Encoding Categorical Variables: One-Hot Encoding, Dummy Coding, Effect Coding, Pros and Cons of Categorical Variable Encodings; Dealing with Large Categorical Variables: Feature Hashing, Bin Counting. **Dimensionality Reduction:** Squashing the Data Pancake with PCA: Intuition, Derivation: Linear Projection, Variance and Empirical Variance, Principal Components: First Formulation, Principal Components: Matrix-Vector Formulation, General Solution of the Principal Components; Transforming Features, Implementing PCA: PCA in Action, Whitening and ZCA, Considerations and Limitations of PCA	12Hrs
UNIT – 4	**Nonlinear Featurization via K-Means Model Stacking:** k-Means Clustering, Clustering as Surface Tiling, k-Means Featurization for Classification: Alternative Dense Featurization, Pros, Cons, and Gotchas	12Hrs
UNIT – 5	**Item-Based Collaborative Filtering, First Pass:** Data Import, Cleaning, and Feature Parsing, Academic Paper Recommender: Naive Approach, Second Pass: More Engineering and a Smarter Model, Academic Paper Recommender: Take 2, Third Pass: More Features is More Information, Academic Paper Recommender: Take 3	12Hrs
	Total	58Hrs

*Note:

Text Books:

1. "Feature Engineering for Machine Learning Principles and Techniques for Data Scientists", Alice Zheng& Amanda Casari, O'REILLY, 2018
2. "Feature Engineering and Selection: A Practical Approach for Predictive Models", Max Kuhn, Kjell Johnson, CRC Press, 2019

I Semester	CRYPTOGRAPHY & NETWORK SECURITY (PROGRAM ELECTIVE-III)	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES:

- Explain the objectives of information security
- Explain the importance and application of each of confidentiality, integrity, authentication and availability
- Understand the basic categories of threats to computers and networks
- Discusses the Mathematics of Cryptography
- Discuss the fundamental ideas of Symmetric and Asymmetric Cryptographic Algorithms
- Discusses the Network layer, Transport Layer and Application Layer Protocols Enhanced security mechanisms

Course Outcomes: At the end of the course, student will be able to

CO	Course Outcomes	Knowledge Level (K)#
CO1	Understand security issues related to computer networks and learn different symmetric key techniques	K2
CO2	Apply mathematic of cryptography for symmetric and Asymmetric algorithms and apply this knowledge to understand the Cryptographic algorithms	K3
CO3	Describe different types of symmetric and Asymmetric algorithms	K2
CO4	Apply algorithms of Hash functions, message authentication and digital signature and their importance to the security	K3
CO5	Explore enhanced security protocols of Application Layer, Transport Layer and Network layer	K6

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	2	2	1
CO2	3	1	3	3	3	2
CO3	2	1	3	2	2	1
CO4	3	1	3	3	3	2
CO5	3	1	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L-1, M-2, H-3)

UNIT	CONTENTS	**Contact Hours**
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. Classical Encryption Techniques-symmetric cipher model, Substitution techniques, Transposition techniques, Rotor Machines, Stegnography.	10Hrs
UNIT – 2	**Introduction to Symmetric Cryptography: Algebraic Structures**-Groups, Rings, Fields, GF(2^n) fields, Polynomials.**Mathematics of Asymmetric cryptography:** Primes, Checking For Primness, Eulers phi-functions, Fermat's Little Theorem, Euler's Theorem, Generating Primes, Primality Testing, Factorization, Chinese Remainder Theorem, Quadratic Congruence, Exponentiation And Logarithm.	12Hrs
UNIT – 3	**Symmetric key Ciphers:** Block Cipher principles, DES, AES, Blowfish, IDEA, Block cipher operation, Stream ciphers: RC4, RC5 **Asymmetric key Ciphers:** Principles of public key cryptosystems, RSA algorithm, Diffie-Hellman Key Exchange, Elgamal Cryptographic system, Elliptic Curve Arithmetic, Elliptic Curve Cryptography.	12Hrs
UNIT – 4	**Cryptographic Hash Functions:** Applications of Cryptographic Hash Functions, Two Simple Hash Functions, Requirements and Security, Hash Functions Based on Cipher Block Chaining, Secure Hash Algorithms (SHA) **Message Authentication Codes:** Message Authentication Requirements, Message Authentication Functions, Requirements for Message Authentication Codes, Security of MAC'S, MAC'S Based On Hash Functions: HMAC, MAC'S Based On Block Ciphers: DAA And CMAC **Digital Signatures:** Digital Signatures, Elgamal Digital Signature Scheme, Elliptic Curve Digital Signature Algorithm, RSA-PSS Digital Signature Algorithm.	12Hrs
UNIT – 5	**Network and Internet Security: Transport-Level Security:** Web Security Considerations, Transport Level Security, HTTPS, SSH. **IP Security:** IP Security Overview, IP Security Policy, Encapsulating Security Payload, Authentication Header Protocol. **Electronic-Mail Security:** Internet-mail Security, Email Format, Email Threats and Comprehensive Email Security, S/MIME, PGP.	12Hrs
	Total	58Hrs

*Note:

TEXT BOOKS:

1. Cryptography and Network Security - Principles and Practice: William Stallings, Pearson Education, 7th Edition, 2017
   Cryptography and Network Security: Behrouz A. Forouzan Debdeep, Mc Graw Hill

Reference Books:

1. Charlie Kaufman (2002), Network Security: Private Communication in a Public World, 2nd edition, Prentice Hall of India, New Delhi.
2. Atul Kahate (2008), Cryptography and Network Security, 2nd edition, Tata Mc Grawhill, India.
   Robert Bragg, Mark Rhodes (2004), Network Security: The complete reference, Tata Mc Grawhill, India.

I Semester	INFORMATION RETRIEVAL SYSTEMS	L	T	P	C
		3	0	0	3

Course Outcomes:

		Knowledge Level (K)#
CO1	Understand the fundamentals of Information Retrieval (IR) systems, including domain analysis and system evaluation	K2
CO2	Apply appropriate data structures and algorithms for indexing and retrieval in IR systems	K3
CO3	Analyze various text indexing techniques like Inverted Files, Signature Files, and PAT Trees	K4
CO4	Evaluate stemming algorithms and methods for thesaurus construction in IR systems	K5
CO5	Implement and compare different string searching algorithms used in IR	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	2	2	1
CO2	3	1	3	3	3	2
CO3	3	1	3	3	3	2
CO4	3	1	3	3	3	2
CO5	3	1	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Introduction to Information storage and retrieval systems:** Domain Analysis of IR systems, IR and other types of Information Systems, IR System Evaluation **Introduction to Data structures and algorithms related to Information Retrieval**: Basic Concepts, Data structures, Algorithms.	10Hrs
UNIT – 2	**Inverted Files and Signature Files:** Introduction, Structures used in Inverted Files, Building an Inverted file using a sorted array, Modifications to the Basic Techniques. **Signature Files:** Concepts of Signature files, Compression, Vertical Partitioning, Horizontal Partitioning.	12Hrs

UNIT – 3	**New Indices for Text, Lexical Analysis and Stoplists: PAT Trees and PAT Arrays:** Introduction, PAT Tree structure, Algorithms on the PAT Trees, Building PAT Trees as PATRICA Trees, PAT representation as Arrays. Stoplists.	12Hrs
UNIT – 4	**Stemming Algorithms and Thesaurus Construction:** Types of Stemming algorithms, Experimental Evaluations of Stemming, Stemming to Compress Inverted Files. **Thesaurus Construction:** Features of Thesauri, Thesaurus Construction, Thesaurus construction from Texts, Merging existing Thesauri.	12Hrs
UNIT – 5	**String Searching Algorithms:** Introduction, Preliminaries, The Naive Algorithm, The Knutt-Morris-Pratt Algorithm, The Boyer-Moore Algorithm, The Shift-Or Algorithm, The Karp-Rabin Algorithm.	12Hrs
	Total	58Hrs

*Note:

TEXT BOOKS

1. Modern Information Retrieval,Ricardo Baeza-Yates, Neto, PEA,2007.
2. Information Storage and Retrieval Systems: Theory and Implementation, Kowalski, Gerald, Mark Academic Press, 2000.

I Semester	NATURAL LANGUAGE PROCESSING	L	T	P	C
		3	0	0	3

Course Objectives: This course introduces the fundamental concepts and techniques of natural language processing (NLP).

- Students will gain an in-depth understanding of the computational properties of natural languages and the commonly used algorithms for processing linguistic information.
- The course examines NLP models and algorithms using both the traditional symbolic and the more recent statistical approaches.
- Enable students to be capable to describe the application based on natural language processing and to show the points of syntactic, semantic and pragmatic processing.

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Demonstrate a given text with basic Language features	K3
CO2	Design an innovative application using NLP components	K6
CO3	Explain a rule based system to tackle morphology/syntax of a language	K2
CO4	Design a tag set to be used for statistical processing for real-time applications	K6
C05	Compare and contrast the use of different statistical approaches for different types of NLP applications	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	2	2
CO2	3	2	3	3	3	3
CO3	2	1	3	3	2	1
CO4	3	2	3	3	3	2
CO5	3	1	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**INTRODUCTION:** Origins and challenges of NLP – Language Modeling: Grammar-based LM, Statistical LM – Regular Expressions, Finite-State Automata – English Morphology, Transducers for lexicon and rules, Tokenization, Detecting and Correcting Spelling Errors, Minimum Edit Distance	10Hrs

UNIT – 2	**WORD LEVEL ANALYSIS:** Unsmoothed N-grams, Evaluating N-grams, Smoothing, Interpolation and Backoff – Word Classes, Part- of-Speech Tagging, Rule-based, Stochastic and Transformation-based tagging, Issues in PoS tagging – Hidden Markov and Maximum Entropy models	12Hrs
UNIT – 3	**SYNTACTIC ANALYSIS**: Context-Free Grammars, Grammar rules for English, Treebanks, Normal Forms for grammar – Dependency Grammar – Syntactic Parsing, Ambiguity, Dynamic Programming parsing – Shallow parsing Probabilistic CFG, Probabilistic CYK, Probabilistic Lexicalized CFGs – Feature structures, Unification of feature structures	12Hrs
UNIT – 4	**SEMANTICS AND PRAGMATICS:** Requirements for representation, First-Order Logic, Description Logics – Syntax-Driven Semantic analysis, Semantic attachments – Word Senses, Relations between Senses, Thematic Roles, selectional restrictions – Word Sense Disambiguation, WSD using Supervised, Dictionary & Thesaurus, Bootstrapping methods – Word Similarity using Thesaurus and Distributional methods.	12Hrs
UNIT – 5	**DISCOURSE ANALYSIS AND LEXICAL RESOURCES:** Discourse segmentation, Coherence – Reference Phenomena, Anaphora Resolution using Hobbs and Centering Algorithm – Coreference Resolution – Resources: Porter Stemmer, Lemmatizer, Penn Treebank, Brill's Tagger, WordNet, PropBank, FrameNet, Brown Corpus, British National Corpus (BNC	12Hrs
	Total	58Hrs

*Note:

Text Books:

3. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech, 2nd Edition, Daniel Jurafsky, James H. Martin - Pearson Publication,2014.
4. Natural Language Processing with Python, First Edition, Steven Bird, Ewan Klein and Edward Loper, OReilly Media,2009.

Reference Books:

4. Language Processing with Java and Ling Pipe Cookbook, 1st Edition, Breck Baldwin, Atlantic Publisher, 2015.
5. Natural Language Processing with Java, 2nd Edition, Richard M Reese, OReilly Media,2015.
6. Handbook of Natural Language Processing, Second, Nitin Indurkhya and Fred J. Damerau, Chapman and Hall/CRC Press, 2010.Edition

   Natural Language Processing and Information Retrieval, 3rd Edition, Tanveer Siddiqui, U.S. Tiwary, Oxford University Press,2008.

I Semester	DATA ENGINEERING	L	T	P	C
		3	0	0	3

Course Outcomes:

		Knowledge Level (K)#
CO1	Understand Data Engineering Life cycle	K2
CO2	Apply appropriate data modeling techniques for different types of data. (L3).	K3
CO3	Evaluate and select appropriate technologies and frameworks for specific data engineering tasks	K5
CO4	Implement data quality checks and governance processes to ensure data reliability and compliance.	K3

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	2	1
CO2	3	1	3	3	3	2
CO3	2	2	3	3	3	2
CO4	3	2	3	3	3	3

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Introduction to Data Engineering:** Definition, Data Engineering Life Cycle, Evolution of Data Engineer, Data Engineering Versus Data Science, Data Engineering Skills and Activities, Data Maturity, Data Maturity Model, Skills of a Data Engineer, Business Responsibilities, Technical Responsibilities, Data Engineers and Other Technical Roles	10Hrs
UNIT – 2	**Data Engineering Life Cycle:** Data Life Cycle Versus Data Engineering Life Cycle, Generation: Source System, Storage, Ingestion, Transformation, Serving Data. Major undercurrents across the Data Engineering Life Cycle: Security, Data Management, Data Ops, Data Architecture, Orchestration, Software Engineering.	12Hrs
UNIT – 3	**Designing Good Data Architecture**: Enterprise Architecture, Data Architecture, Principles of Good Data Architecture, Major Architecture Concepts. Data Generation in Source Systems: Sources of Data, Files and Unstructured Data, APIs, Application Databases (OLTP), OLAP, Change Data Capture, Logs, Database Logs, CRUD, Source System Practical Details.	12Hrs

UNIT – 4	**Storage:** Raw Ingredients of Data Storage, Data Storage Systems, Data Engineering Storage Abstractions, Data warehouse, Data Lake, Data Lakehouse. Ingestion: Data Ingestion, Key Engineering considerations for the Ingestion Phase, Batch Ingestion Considerations, Message and Stream Ingestion Considerations, Ways to Ing	12Hrs
UNIT – 5	**Queries, Modeling and Transformation**: Queries, Life of a Query, Query Optimizer, Queries on Streaming Data, Data Modelling, Modeling Streaming Data, Transformations, Streaming Transformations and Processing. Serving Data for Analytics, Machine Learning and Reverse ETL: General Considerations for serving Data, Business Analytics, Operational Analytics, Embedded Analytics, Ways to serve data for analytics and ML, Reverse ETL.	12Hrs
	Total	58Hrs

*Note:

Textbooks:

1. Joe Reis, Matt Housley, Fundamentals of Data Engineering, O'Reilly Media, Inc.,June 2022,ISBN: 9781098108304

Reference Books:

1. Paul Crickard , Data Engineering with Python,Packt Publishing, October 2020.
2. Ralph Kimball, Margy Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Wiley, 3rd Edition, 2013
3. James Densmore, Data Pipelines Pocket Reference: Moving and Processing Data for Analytics, O'Reilly Media, 1st Edition, 2021

I Semester	ADVANCED ALGORITHMS ANALYSIS LAB	L	T	P	C
		0	1	2	2

List of Experiments:

UNIT	CONTENTS	Contact Hours
Experiment– 1	Implement assignment problem using Brute Force method	3Hrs
Experiment– 2	Perform multiplication of long integers using divide and conquer method.	3Hrs
Experiment– 3	Implement a solution for the knapsack problem using the Greedy method	3Hrs
Experiment– 4	Implement Gaussian elimination method.	
Experiment– 5	Implement LU decomposition	3Hrs
Experiment– 6	Implement Warshall algorithm	3Hrs
Experiment– 7	Implement the Rabin Karp algorithm.	3Hrs
Experiment– 8	Implement the KMP algorithm.	3Hrs
Experiment– 9	Implement Harspool algorithm	3Hrs
Experiment– 10	Implement max-flow problem.	3Hrs
	Total	30Hrs

Text Book:

1. Design and Analysis of Algorithms, S.Sridhar, OXFORD University Press

References Books

1. Introduction to Algorithms, second edition, T.H. Cormen, C.E. Leiserson, R.L. Rivest and C. Stein, PHI Pvt. Ltd./ Pearson Education.
2. Fundamentals of Computer Algorithms, Ellis Horowitz, Satraj Sahni and Rajasekharam, Universities Press.
3. Design and Analysis of algorithms, Aho, Ullman and Hopcroft, Pearson education

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

R25 M.TECH DATA SCIENCE COURSE STRUCTURE AND SYLLABUS

I Semester	DATA SCIENCE APPLICATIONS USING R/PYTHON LAB	L	T	P	C
		0	1	2	2

List of Experiments:

UNIT	CONTENTS	**Contact Hours**
Experiment– 1	Write a Python Program to Find the Sum of the Series: 1 + 1/2 + 1/3 + .. + 1/N	3Hrs
Experiment– 2	Write a Python Program to Split the array and add the first part to the end	3Hrs
Experiment– 3	Write a Python Program to Create a List of Tuples with the First Element as the Number and Second Element as the Square of the Number	3Hrs
Experiment– 4	Write a Python program to count number of vowels using sets in given string	3Hrs
Experiment– 5	Write a program to implement permutation of a given string using inbuilt function**.**	3Hrs
Experiment– 6	Write a python program to sort list of dictionaries by values in Python – Using lambda function	3Hrs
Experiment– 7	Write a Python Program for following sorting: i. Quick Sort ii. Heap Sort	3Hrs
Experiment– 8	Write a Python Program to Reverse a String Using Recursion	3Hrs
Experiment– 9	Write a Python Program to Count the Number of Words in a Text File	3Hrs
Experiment– 10	Write a Python Program to Read the Contents of a File in Reverse Order	3Hrs
Experiment– 11	Write a program to Merge and Join DataFrames with Pandas in Python	3Hrs
Experiment– 12	Write a program to implement Merge and Join DataFrames with Python Pandas	3Hrs
Experiment– 13	Write a Python Program to Append the Contents of One File to Another File	3Hrs
Experiment– 14	How to install and Load CSV files to Python Pandas	3Hrs
Experiment– 15	Write a program to implement Data analysis and Visualization with Python using pandas	3Hrs
Experiment– 16	Write a program to Implement Plotting Functions in python pandas.	3Hrs
	Total	48Hrs

II Semester	LARGE LANGUAGE MODELS	L	T	P	C
		3	1	0	4

Course Outcomes:

		Knowledge Level (K)#
CO1	Explore the concept of Language Modeling Techniques and Implement the AND,OR and XOR logic	K3
CO2	Discuss the different types of Word Embedding	K2
CO3	Explain the concept transformer decoding blocks	K2
CO4	Discuss the concept of Pretraining Applications and Limitations	K4
CO5	Demonstrate the concept of Handling Memory for LLM Agents	K3

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	3	2
CO2	2	1	3	2	2	1
CO3	2	1	3	3	3	2
CO4	3	2	3	3	3	2
CO5	3	2	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT -1	**Introduction**- What is a Language Model?, Evolution of Language Modelling Technologies, Scaling Laws in Language Models, Evolution of LLMs, The Emergence and Development of LLMs, Implications of Encoder-Decoder in LLM Development, Optimising Scale and Resource Efficiency in LLMs **Neural Networks**: The Perceptron, Definition, Implementing AND, OR, and XOR Logic, Multilayer Perceptron, Neural Networks, Types of Activation Functions.	10Hrs
UNIT – 2	**Word Embedding**-Distributional Hypothesis, Vector Semantics, Defining and Measuring Semantic Similarity, Types of Word Embedding, Frequency-Based Embeddings, Word2Vec, Global Vectors for Word Representation, Fast Text, Bias in Word Embedding, Limitations of Word Embedding Methods, Applications of Word Embeddings.	12Hrs

UNIT -3	**Transformers-** Self-Attention, Multi-Head Self-Attention, Transformer Encoder Block, Components of the Transformer Encoder Block, Feed-Forward Neural Network, Layer Normalisation, Residual Connections, Transformer Decoder Block, Masked Multi-Head Self-Attention, Cross-Attention (Encoder-Decoder Attention), Positional Embeddings, Types of Positional Embeddings, Rotary Position Embedding, Efficient Attention Mechanisms,	12Hrs
UNIT -4	**Language Model Pretraining-** Embeddings from Language Model, Architecture and Training of ELMo, Applications of ELMo, Limitations of ELMo, Evaluation Datasets, Encoder-Based Pretraining, Fundamentals of Encoder-Based Models, Training Paradigm, BERT Pretraining, Applications and Limitations, Decoder-Based Pretraining, Decoder-Based Architecture, Training Paradigm,GPT Pretraining, Applications and Limitations, Encoder-Decoder Based Pretraining, Architecture, Joint Pretraining Strategy,T5 Pretraining, Applications and Limitations, Emergence of Large Language Models, Limitations of Pretraining.	
UNIT -5	**Augmented Large Language Models- Retrieval-Augmented Generation**, Indexing in RAGs, Context Searching in RAGs, Prompting in RAGs, Inferencing in RAGs, Comparison of RAGs with LLMs, Evaluation of RAGs, Assessing of Retrieval Quality, Generation Quality, Knowledge Integration and Factuality Evaluation, Response Time and Efficiency, User Satisfaction, RAGAs Framework for RAG Evaluation, Tool Calling with LLMs, Autonomously Determining Which Tools to Use and Where, Examples of Different Tools, Evaluation of Code Generation Capabilities of Agents, Error Handling and Optimisation, LLM Augmentation with Agents, Reasoning in LLM Agents, Planning in LLM Agents, Handling Memory in LLM Agents	12Hrs

Text Books:

1. “Introduction to Large Language Models” Generative AI for Text by Tanmoy Chakrborty-25 December 2024, Wiley Publications.

II Semester	MACHINE LEARNING	L	T	P	C
		3	1	0	4

Course Outcomes:

		Knowledge Level (K)#
CO1	Enumerate the Fundamentals of Machine Learning	K1
CO2	Build Nearest neighbour based models	K3
CO3	Apply Models based on decision trees and Bayes rule	K3
CO4	Choose appropriate clustering technique	K4
CO5	Determine algorithms to generate code for a target machine	K5

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	1	3	2	2	1
CO2	2	1	3	3	3	2
CO3	3	2	3	3	3	2
CO4	3	2	3	3	3	2
CO5	3	2	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT -1	**Introduction to Machine Learning:** Evolution of Machine Learning, Paradigms for ML, Learning by Rote, Learning by Induction, Reinforcement Learning, Types of Data, Matching, Stages in Machine Learning, Data Acquisition, Feature Engineering, Data Representation, Model Selection, Model Learning, Model Evaluation, Model Prediction, Search and Learning, Data Sets.	10Hrs
UNIT – 2	**Nearest Neighbor-Based Models:** Introduction to Proximity Measures, Distance Measures, Non-Metric Similarity Functions, Proximity Between Binary Patterns, Different Classification Algorithms Based on the Distance Measures ,K-Nearest Neighbor Classifier, Radius Distance Nearest Neighbor Algorithm, KNN Regression, Performance of Classifiers, Performance of Regression Algorithms.	12Hrs

UNIT -3	**Models Based on Decision Trees**: Decision Trees for Classification, Impurity Measures, Properties, Regression Based on Decision Trees, Bias–Variance Trade-off, Random Forests for Classification and Regression.**The Bayes Classifier:** Introduction to the Bayes Classifier, Bayes' Rule and Inference, The Bayes Classifier and its Optimality, Multi-Class Classification | Class Conditional Independence and Naive Bayes Classifier (NBC)	12Hrs
UNIT -4	**Linear Discriminants for Machine Learning**: Introduction to Linear Discriminants, Linear Discriminants for Classification, Perceptron Classifier, Perceptron Learning Algorithm, Support Vector Machines, Linearly Non-Separable Case, Non-linear SVM, Kernel Trick, Logistic Regression, Linear Regression, Multi-Layer Perceptrons (MLPs), Backpropagation for Training an MLP.	
UNIT -5	**Clustering** : Introduction to Clustering, Partitioning of Data, Matrix Factorization | Clustering of Patterns, Divisive Clustering, Agglomerative Clustering, Partitional Clustering, K-Means Clustering, Soft Partitioning, Soft Clustering, Fuzzy C-Means Clustering, Rough Clustering, Rough K-Means Clustering Algorithm, Expectation Maximization-Based Clustering, Spectral Clustering.	12Hrs

Text Books:

2. "Machine Learning Theory and Practice", M N Murthy, V S Ananthanarayana, Universities Press (India), 2024

Reference Books:

1. "Machine Learning", Tom M. Mitchell, McGraw-Hill Publication, 2017
2. "Machine Learning in Action",Peter Harrington, DreamTech
3. "Introduction to Data Mining", Pang-Ning Tan, Michel Stenbach, Vipin Kumar, 7th Edition, 2019.

II Semester	FOUNDATIONS OF DATA ANALYTICS	L	T	P	C
		3	1	0	4

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Understand the impact of data analytics for business decisions and strategy Carry out data analysis/statistical analysis	K1
CO2	Discover interesting patterns, analyze supervised and unsupervised and estimate the accuracy of the algorithms.	K2
CO3	Understand the various search methods and visualization techniques	K1
CO4	Determine algorithms to generate code for a target machine	K3

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	1	3	2	2	1
CO2	2	1	3	3	3	2
CO3	3	1	3	3	3	2
CO4	3	2	3	3	3	2
CO5	3	2	3	3	3	2

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**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	10Hrs
UNIT – 2	**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.	12Hrs
UNIT – 3	**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.	12Hrs

UNIT – 4	**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	12Hrs
UNIT – 5	**V Data Visualization**: Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data and Relations.	12Hrs
	Total	58Hrs

*Note:

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:

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

II Semester	DEEP LEARNING	L	T	P	C
		3	0	0	3

Course Objectives:

- The objective of this course is to cover the fundamentals of neural networks as well as some advanced topics such as recurrent neural networks, long short-term memory cells and convolution neural networks.

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Understand basic concepts of biological and computational neurons, perceptron models, and learning algorithms.	K2
CO2	Apply feedforward neural networks and optimization algorithms like backpropagation, gradient descent, and auto encoders.	K3
CO3	Analyze and evaluate modern optimization methods and regularization techniques to train better deep networks	K4
CO4	Apply advanced neural network architectures such as RNNs, LSTMs, CNNs, and RBMs to sequence and image data	K3
CO5	Evaluate and compare recent deep learning trends including transformers, GPTs, and their applications in NLP, vision, and speech	K5

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	2	2	1	1
CO2	3	2	3	3	3	2
CO3	3	2	3	3	3	2
CO4	3	2	3	3	3	2
CO5	3	2	2	3	3	3

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT-1	**Basics-** Biological Neuron, Idea of computational units, McCulloch–Pitts unit and Thresholding logic, Linear Perceptron, Perceptron Learning Algorithm, Linear separability, Convergence theorem for Perceptron Learning Algorithm.	10Hrs

UNIT-2	**Feed forward Networks**-Multilayer Perceptron, Gradient Descent, Backpropagation, Empirical Risk Minimization, regularization, auto encoders. Deep Neural Networks: Difficulty of training deep neural networks, Greedy layer wise training	12Hrs
UNIT-3	**Better Training of Neural Networks-**Newer optimization methods for neural networks (Adagrad, adadelta, rmsprop, adam, NAG), second order methods for training, Saddle point problem in neural networks, Regularization methods (dropout, drop connect, batch normalization).	12Hrs
UNIT-4	**Recurrent Neural Networks-** Back propagation through time, Long Short-Term Memory, Gated Recurrent Units, Bidirectional LSTMs, Bidirectional RNNs. Convolutional Neural Networks: LeNet, AlexNet. Generative models: Restrictive Boltzmann Machines (RBMs), Introduction to MCMC and Gibbs Sampling, gradient computations in RBMs, Deep Boltzmann Machines.	
UNIT-5	**Recent trends**- Variational Autoencoders, Transformers, GPT Applications: Vision, NLP, Speech	12Hrs

Text Books:

1. Deep Learning, Ian Good fellow and Yoshua Bengio and Aaron Courville, MIT Press, 2016

Reference Books:

1. Neural Networks: A Systematic Introduction, Raúl Rojas,1996
2. Pattern Recognition and Machine Learning, Christopher Bishop,2007

II Semester	DATABASE AND WEB APPLICATION SECURITY	L	T	P	C
		3	0	0	3

Course Objectives:

1. To acquire knowledge on standard algorithms used to provide confidentiality,
2. integrity and authenticity.
3. To design security applications in the field of Information technology.
4. To understand the fundamentals of database design, DB security and SQL
5. extensions to security.
6. To learn the basic concepts of Penetration testing.

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Explain the fundamental concepts of database security, threats, access control models, and inference attacks	K2
CO2	Apply access control mechanisms and analyze SQL injection vulnerabilities in relational and distributed databases	K3
CO3	Evaluate security issues in web applications and implement security measures based on OWASp standards	K4
CO4	Analyze mobile device security threats and apply software and hardware-centric security techniques.	K4
CO5	Apply penetration testing methods and tools to identify and mitigate application and system vulnerabilities.	K3

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	1			1
CO2	3	3	2	2	2	
CO3	3	3	3	3	3	1
CO4	2	2	2	3	3	
CO5	2	3	3	3	3	1

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT-1	**Database Security-**Introduction includes threats, vulnerabilities and breaches, Basics of database design, DB security, concepts, approaches and challenges, types of access controls, Oracle VPD, Discretionary and Mandatory access control: Principles, applications and poly instantiation, Database inference problem, types of inference attacks, distributed database, security levels, SQL-injection: types and advanced concepts	10Hrs

UNIT-2	**Relational Data Model-** Security in relational data model, concurrency controls and locking, SQL extensions to security (oracle as an example), System R concepts, Context and control based access control, Hippocratic databases, Database watermarking, Database intrusion, Secure data outsourcing.	12Hrs
UNIT-3	**Web application security-**Basic principles and concepts, Authentication, Authorization, Browser security principles; XSS and CSRF, same origin policies, File security principles, Secure development and deployment methodologies, Web DB principles, OWASP – Top 10 -Detailed treatment, IoT security.	12Hrs
UNIT-4	**Mobile device security:** Introduction, attack vector and models, hardware centric security aspects, SMS / MMS vulnerabilities, software centric security aspects, mobile web browser security, Application security: Concepts, CIA Triad, Hexad, types of cyber-attacks, Introduction to software development vulnerabilities, code analyzers– Static and dynamic analyzers.	
UNIT-5	**Penetration testing-** Principles and concepts, PT work flows and examples, blind tests, ethical hacking techniques, synthetic transactions, interface testing and fuzzing, SDLC phases and security mandates.	12Hrs

Text Books:

1. Web Application Security, 1st Edition, A Beginners Guide, Bryan and Vincent, McGraw-Hill, 2011
2. Database Security, 1st Edition, Alfred Basta, Melissa Zgola, Course Technology, 2012

Reference Books:

1. Handbook of Database Security: Applications and Trends, Michael Gertz and Sushil Jajodia, Springer, 2008
2. Database and Applications Security, 1st Edition, Integrating Information Security and Data Management, Bhavani Thuraisingham, Auerbach Publications, 2005

II Semester	SOFTWARE PROJECT MANAGEMENT	L	T	P	C
		3	0	0	3

Course Objectives:

At the end of the course, the student shall be able to:

- To describe and determine the purpose and importance of project management from theperspectives of planning, tracking and completion of project
- To compare and differentiate organization structures and project structures
- To implement a project to manage project schedule, expenses and resources with the application ofsuitable project management tools

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Explain conventional and modern software management principles and improvements in software economics.	K2
CO2	Describe life cycle phases and process artifacts used in software development and management	K2
CO3	Apply iterative planning, work breakdown structures, cost/schedule estimation, and model-based architectures.	K3
CO4	Analyze project organization structures, automation tools, and process instrumentation using metrics.	K4
CO5	Evaluate and apply Agile and DevOps principles, practices, and tools in modern software development.	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	2	2		1
CO2	3	2	2	2	1	
CO3	3	3	3	3	2	
CO4	2	3	3	3	3	
CO5	2	2	3	3	2	1
CO6						

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Conventional Software Management:** The waterfall model, conventional software Managementperformance. **Evolution of Software Economics:** Software Economics, pragmatic software cost estimation. **Improving Software Economics:** Reducing Software product size, improving software processes, improving team effectiveness, improving automation, Achieving required quality, peer inspections. **The old way and the new:** The principles of conventional software Engineering, principles of modernsoftware management, transitioning to an iterative process.	10Hrs
UNIT – 2	**Life cycle phases:** Engineering and production stages, inception, Elaboration, construction, transitionphases. **Artifacts of the process:** The artifact sets, Management artifacts, Engineering artifacts, programmaticartifacts.	12Hrs
UNIT – 3	**Model based software architectures:** A Management perspective and technical perspective. **Work Flows of the process:** Software process workflows, Iteration workflows. **Checkpoints of the process: Major** mile stones, Minor Milestones, Periodic status assessments. **Iterative Process Planning:** Work breakdown structures, planning guidelines, cost and scheduleestimating, Iteration planning process, Pragmatic planning.	12Hrs
UNIT – 4	**Project Organizations and Responsibilities:** Line-of-Business Organizations, Project Organizations,evolution of Organizations. **Process Automation:** Automation Building blocks, The Project Environment. **Project Control and Process instrumentation:** The seven core Metrics, Management indicators, quality indicators, life cycle expectations, pragmatic Software Metrics, Metrics automation.	12Hrs
UNIT – 5	Agile Methodology, ADAPTing to Scrum, Patterns for Adopting Scrum, Iterating towards Agility. **Fundamentals of DevOps**: Architecture, Deployments, Orchestration, Need, Instance of applications,DevOps delivery pipeline, DevOps eco system. DevOps adoption in projects: Technology aspects, Agiling capabilities, Tool stack implementation, People aspect, processes	12Hrs
	Total	58Hrs

*Note:

Text Books:

1. Software Project Management, Walker Royce, PEA, 2005.
2. Succeeding with Agile: Software Development Using Scrum, Mike Cohn, Addison Wesley.
3. The DevOps Handbook: How to Create World-Class Agility, Reliability, and Security in Technology Organizations, Gene Kim , John Willis , Patrick Debois , Jez Humb,1st Edition, O'Reilly publications, 2016.

Reference Books:

1. Software Project Management, Bob Hughes,3/e, Mike Cotterell, TMH
2. Software Project Management, Joel Henry, PEA
3. Software Project Management in practice, Pankaj Jalote, PEA, 2005,
4. Effective Software Project Management, Robert K.Wysocki, Wiley,2006.
5. Project Management in IT, Kathy Schwalbe, Cengage

II Semester	BIG DATA ANALYTICS	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES:

1. Optimize business decisions and create competitive advantage with Big Data analytics
2. Introducing Java concepts required for developing map reduce programs
3. Derive business benefit from unstructured data
4. Imparting the architectural concepts of Hadoop and introducing map reduce paradigm
5. To introduce programming tools PIG & HIVE in Hadoop echo system.

Course Outcomes:

		Knowledge Level (K)#
CO1	**Preparing** for data summarization, query, and analysis.	K2
CO2	**Applying** data modelling techniques to large data sets	K3
CO3	**Creating** applications for Big Data analytics.	K2
CO4	**Demonstrate and Develop** the pig and Latin Scripts	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	2	2	3	
CO2	3	3	2	2	3	
CO3	2	3	3	2	3	
CO4	3	2	2	2	3	

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Data structures in Java:** Linked List, Stacks, Queues, Sets, Maps; Generics: Generic classes and Type parameters, Implementing Generic Types, Generic Methods, Wrapper Classes, Concept of Serialization	10Hrs
UNIT – 2	**: Working with Big Data:** Google File System, Hadoop Distributed File System (HDFS) Building blocks of Hadoop (Namenode, Datanode, Secondary Namenode, Job Tracker, Task Tracker), Introducing and Configuring Hadoop cluster (Local, Pseudo-distributed mode, Fully Distributed mode), Configuring XML files.	12Hrs

UNIT – 3	**Writing Map Reduce Programs:** A Weather Dataset, Understanding Hadoop API forMap Reduce Framework (Old and New), Basic programs of Hadoop Map Reduce: Driver code, Mapper code, Reducer code, Record Reader, Combiner, Practitioner	12Hrs
UNIT – 4	**StreamMemory and Spark:** Introduction to Streams Concepts– Stream Data ModelandArchitecture , Stream computing, Sampling Data in a Stream , Filtering Streams ,Counting Distinct Elements in a Stream , Introduction to Spark Concept , Spark Architecture and components , Spark installation , Spark RDD(Resilient Distributed Dataset) – Spark RDD operations	12Hrs
UNIT – 5	Hadoop Programming Made EasierAdmiring the Pig Architecture, Going with the Pig Latin Application Flow, Working through the ABCs of Pig Latin, Evaluating Local and Distributed Modes of Running Pig Scripts, Checking out the Pig Script Interfaces, Scripting with Pig Latin.	12Hrs
	Total	58Hrs

*Note:

TEXT BOOKS:

1. Wiley & Big Java 4th Edition, Cay Horstmann, Wiley John Sons, INC
2. Hadoop: The Definitive Guide by Tom White, 3rd. Edition, O'reilly

REFERENCE BOOKS:

1. Hadoop in Action by Chuck Lam, MANNING Publ.
2. Hadoop for Dummies by Dirk deRoos, Paul C.Zikopoulos, Roman B.Melnyk,Bruce Brown, Rafael Coss
3. Hadoop in Practice by Alex Holmes, MANNING Publ.
4. Big Data Analytics by Dr. A.Krishna Mohan and Dr.E.Laxmi Lydia
5. Hadoop Map Reduce Cookbook, SrinathPerera, ThilinaGunarathne

Software Links:

1. Hadoop:http://hadoop.apache.org/
2. Hive: https://cwiki.apache.org/confluence/display/Hive/Home
3. Piglatin: http://pig.apache.org/docs/r0.7.0/tutorial.html

II Semester	OBJECT ORIENTED ANALYSIS & DESIGN	L	T	P	C
		3	0	0	3

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Describe the structure and inherent complexity of software systems and explain the design of complex systems.	K2
CO2	Apply UML principles and structural modeling techniques to real-world systems using diagrams.	K3
CO3	Construct and interpret class, object, and advanced structural diagrams for software systems.	K3
CO4	Model system behaviors using use cases, interaction, and activity diagrams	K3
C05	Analyze and develop advanced behavioral and architectural UML diagrams including state, component, and deployment diagrams.	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	2	1		
CO2	3	3	3	2	3	
CO3	3	3	3	2	3	1
CO4	2	3	3	2	3	
CO5	2	3	3	3	3	

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Introduction:** The Structure of Complex systems, The Inherent Complexity of Software, Attributes of Complex System, Organized and Disorganized Complexity, Bringing Order to Chaos, Designing Complex Systems. **Case Study:** System Architecture: Satellite-Based Navigation	10Hrs
UNIT – 2	**Introduction to UML:** Importance of modeling, principles of modeling, object oriented modeling, conceptual model of the UML, Architecture, and Software Development Life Cycle. **Basic Structural Modeling:** Classes, Relationships, common Mechanisms, and diagrams. **Case Study:** Control System: Traffic Management.	12Hrs

UNIT – 3	**Class & Object Diagrams:** Terms, concepts, modeling techniques for Class & Object Diagrams. **Advanced Structural Modeling:** Advanced classes, advanced relationships, Interfaces, Types and Roles, Packages. **Case Study:** AI: Cryptanalysis.	12Hrs
UNIT – 4	**Basic Behavioral Modeling-I:** Interactions, Interaction diagrams Use cases, Use case Diagrams, Activity Diagrams. **Case Study:** Web Application: Vacation Tracking System.	12Hrs
UNIT – 5	**Advanced Behavioral Modeling:** Events and signals, state machines, processes and Threads, time and space, state chart diagrams. **Architectural Modeling:** Component, Deployment, Component diagrams and Deployment diagrams **Case Study:** Weather Forecasting	12Hrs
	Total	58Hrs

*Note:

Text Books:

1. Grady BOOCH, Robert A. Maksimchuk, Michael W. ENGLE, Bobbi J. Young, Jim Conallen, Kellia Houston , "Object- Oriented Analysis and Design with Applications", 3rd edition,2013, PEARSON.
2. Grady Booch, James Rumbaugh, Ivar Jacobson: The Unified Modeling Language User Guide,Pearson Education.

Reference Books:

1. Meilir Page-Jones: Fundamentals of Object Oriented Design in UML, Pearson Education.
2. Pascal Roques: Modeling Software Systems Using UML2, WILEY- Dreamtech India Pvt. Ltd.
3. Atul Kahate: Object Oriented Analysis & Design, The McGraw-Hill Companies.

II Semester	QUANTUM COMPUTING	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: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Explain the interdisciplinary foundations of quantum computing from mathematics, physics, and biology	K2
CO2	Apply the principles of quantum mechanics and linear algebra to analyze quantum states and measurements.	K3
CO3	Construct and interpret quantum circuits using single and multi-qubit gates, and represent states using Bloch sphere.	K3
CO4	Analyze and apply quantum algorithms such as Deutsch, Shor, and Grover to relevant computational problems.	K4
C05	Evaluate quantum error correction methods, and compare classical vs quantum information and cryptography techniques.	K4

#Based on suggested Revised BTL

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	2	3	3	1	
CO2	2	1	3	3	2	
CO3	2	1	3	2	3	1
CO4	3	2	3	3	3	
CO5	3	2	3	2	2	2
CO6						

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**History of Quantum Computing:** Importance of Mathematics, Physics and Biology. Introduction to Quantum Computing: Bits Vs Qubits, Classical Vs Quantum logical operations	10Hrs
UNIT – 2	**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	12Hrs

	in bases other than computational basis. Background Biology: Basic concepts of Genomics and Proteomics (Central Dogma)	
UNIT – 3	**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.	12Hrs
UNIT – 4	**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	12Hrs
UNIT – 5	**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	12Hrs
	Total	58Hrs

*Note:

Text Books:

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

II Semester	INFORMATION THEORY AND CODING	L	T	P	C
		3	0	0	3

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Understand the mathematical foundation of information theory and various source coding techniques.	K2
CO2	Analyze and apply linear block codes and syndrome decoding techniques in data storage and transmission systems	K3
CO3	Apply cyclic codes for error detection and correction and evaluate their decoding methods.	K3
CO4	Design convolutional coding schemes and implement Viterbi and sequential decoding for error control.	K4
CO5	Explain and implement BCH codes and decoding procedures for single and double error correction	K3

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	3	1	2
CO2	2	2	3	3	2	2
CO3	2	2	3	3	2	2
CO4	3	2	3	3	3	3
CO5	3	2	3	3	2	3

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	Coding for Reliable Digital Transmission and storage Mathematical model of Information, A Logarithmic Measure of Information, Average and Mutual Information and Entropy, Types of Errors, Error Control Strategies. Source Codes: Shannon-fano coding, Huffman coding	10Hrs
UNIT – 2	**Linear Block Codes:** Introduction to Linear Block Codes, Syndrome and Error Detection, Minimum Distance of a Block code, Error-Detecting and Error-correcting Capabilities of a Block code, Standard array and Syndrome Decoding, Probability of an undetected error for Linear Codes over a BSC, Hamming Codes. Applications of Block codes for Error control in data storage system	12Hrs
UNIT – 3	**Cyclic Codes:** Description, Generator and Parity-check Matrices, Encoding, Syndrome Computation and Error Detection, Decoding,	12Hrs

	Cyclic Hamming Codes, shortened cyclic codes, Error-trapping decoding for cyclic codes, Majority logic decoding for cyclic codes.	
UNIT – 4	**Convolutional Codes:** Encoding of Convolutional Codes- Structural and Distance Properties, state, tree, trellis diagrams, maximum likelihood decoding, Sequential decoding, Majority- logic decoding of Convolution codes. Application of Viterbi Decoding and Sequential Decoding, Applications of Convolutional codes in ARQ system.	12Hrs
UNIT – 5	**BCH Codes:** Minimum distance and BCH bounds, Decoding procedure for BCH codes, Syndrome computation and iterative algorithms, Error locations polynomials for single and double error correction	12Hrs
	Total	58Hrs

*Note:

TEXT BOOKS:

1. Error Control Coding- Fundamentals and Applications –Shu Lin, Daniel J.Costello,Jr, Prentice Hall, Inc 2014. 2. Error Correcting Coding Theory-Man Young Rhee, McGraw – Hill Publishing 1989

II Semester	BLOCK CHAIN TECHNOLOGIES	L	T	P	C
		3	0	0	3

Course Objectives:

- To learn the fundamentals of Block Chain and various types of block chain and consensus mechanism.
- To understand public block chain system, Private block chain system and consortium block chain.
- Able to know the security issues of blockchain technology.

Course Outcomes: At the end of the course, student will be able to (Four to Six)

		Knowledge Level (K)#
CO1	Understand the fundamentals of blockchain technology, types of blockchains, consensus mechanisms, and cryptocurrencies.	K2
CO2	Analyze public blockchain systems and smart contracts with use cases in Ethereum and industry.	K4
CO3	Compare private and consortium blockchains and apply key concepts in platforms like Hyperledger and Corda.	K3
CO4	Examine security, privacy, performance, and compliance issues in blockchain networks.	K4
CO5	Develop and test blockchain applications using Python and Hyperledger Fabric through case studies	K3

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6
CO1	2	1	3	2	2	2
CO2	2	2	3	2	2	2
CO3	2	2	3	2	3	2
CO4	3	2	3	2	2	3
CO5	3	2	3	2	3	3

(Please fill the above with Levels of Correlation, viz., L, M, H)

UNIT	CONTENTS	**Contact Hours**
UNIT – 1	**Fundamentals of Blockchain:** Introduction, Origin of Blockchain, Blockchain Solution, Components of Blockchain, Block in a Blockchain, The Technology and the Future. Blockchain Types and Consensus Mechanism: Introduction, Decentralization and Distribution, Types of Blockchain, Consensus Protocol.	10Hrs

	Cryptocurrency: Bitcoin, Altcoin and Token: Introduction, Bitcoin and the Cryptocurrency, Cryptocurrency Basics, Types of Cryptocurrencies, Cryptocurrency Usage.	
UNIT – 2	**Public Blockchain System:** Introduction, Public Blockchain, Popular Public Blockchains, The Bitcoin Blockchain, Ethereum Blockchain. Smart Contracts: Introduction, Smart Contract, Characteristics of a Smart Contract, Types of Smart Contracts, Types of Oracles, Smart Contracts in Ethereum, Smart Contracts in Industry.	12Hrs
UNIT – 3	**Private Blockchain System:** Introduction, Key Characteristics of Private Blockchain, Private Blockchain, Private Blockchain Examples, Private Blockchain and Open Source, E- commerce Site Example, Various Commands (Instructions) in E-commerce Blockchain, Smart Contract in Private Environment, State Machine, Different Algorithms of Permissioned Blockchain, Byzantine Fault, Multichain. Consortium Blockchain: Introduction, Key Characteristics of Consortium Blockchain, Need of Consortium Blockchain, Hyperledger Platform, Overview of Ripple, Overview of Corda.	12Hrs
UNIT – 4	**Security in Blockchain:** Introduction, Security Aspects in Bitcoin, Security and Privacy Challenges of Blockchain in General, Performance and Scalability, Identity Management and Authentication, Regulatory Compliance and Assurance, Safeguarding Blockchain Smart Contract (DApp), Security Aspects in Hyperledger Fabric.	12Hrs
UNIT – 5	**Blockchain Case Studies:** Case Study 1 – Retail, Case Study 2 – Banking and Financial Services, Case Study 3 – Healthcare, Case Study 4 – Energy and Utilities. **Blockchain Platform using Python**: Introduction, Learn How to Use Python Online Editor, Basic Programming Using Python, Python Packages for Blockchain. Blockchain platform using Hyperledger Fabric: Introduction, Components of Hyperledger Fabric Network, Chain codes from Developer.ibm.com, Blockchain Application Using Fabric Java SDK.	12Hrs
	Total	58Hrs

Text book:

1. “Block chain Technology”, Chandramouli Subramanian, Asha A.George, Abhilasj K A and Meena Karthikeyan , Universities Press.

Reference Books:

1. Blockchain Blue print for Economy, Melanie Swan, SPD Oreilly.
2. Blockchain for Business, Jai Singh Arun, Jerry Cuomo, Nitin Gauar, Pearson Addition Wesley

II Semester	LARGE LANGUAGE MODELS LAB	L	T	P	C
		0	1	2	2

Software Required: Python

UNIT	CONTENTS	**Contact Hours**
Experiment– 1	Write a program to implement word Tokenizer, Sentence and Paragraph Tokenizers. Check how many words are there in any corpus. Also check how many distinct words are there?	3Hrs
Experiment– 2	Write a python program to eliminate stopwords using nltk	3Hrs
Experiment– 3	Write a python program to perform tokenization by word and sentence using nltk	3Hrs
Experiment– 4	Write a program to implement Perceptron	3Hrs
Experiment– 5	Write a Program to implement AND OR gates using Perceptron	3Hrs
Experiment– 6	Applying the Autoencoder algorithms for encoding the real-world data	3Hrs
Experiment– 7	Word Analysis	3Hrs
Experiment– 8	Word Generation	3Hrs
Experiment– 9	Building Chunker	3Hrs
Experiment– 10	Building POS Tagger	3Hrs
Experiment– 11	POS Tagging: Hidden Markov Model	3Hrs
Experiment– 12	POS Tagging: Viterbi Decoding	3Hrs
Experiment– 15	Demonstrate the use of Expectation Maximization based clustering algorithm	3Hrs

II Semester	MACHINE LEARNING LAB	L	T	P	C
		0	1	2	2

Course Objectives:

- To learn about computing central tendency measures and Data preprocessing techniques
- To learn about classification and regression algorithms
- To apply different clustering algorithms for a problem.

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

		Knowledge Level (K)#
CO1	Develop program for computing central tendency measures and Apply Data Preprocessing techniques	
CO2	Build Classifiers using KNN, Decision Tree, Random Forest algorithms	
CO3	Implement classification algorithms such as Naïve Bayes, SVM, Multi-Layer Perceptron	
CO4	Apply clustering algorithms such as K-Means, Fuzzy C-Means and Expectation Maximization for a problem	

Software Required: Python/R/Weka

UNIT	CONTENTS	**Contact Hours**
Experiment– 1	Compute Central Tendency Measures: Mean, Median, Mode Measure of Dispersion: Variance, Standard Deviation.	3Hrs
Experiment– 2	Apply the following Pre-processing techniques for a given dataset. a. Attribute selection b. Handling Missing Values c. Discretization d. Elimination of Outliers	3Hrs
Experiment– 3	Apply KNN algorithm for classification and regression	3Hrs
Experiment– 4	Demonstrate decision tree algorithm for a classification problem and perform parameter tuning for better results	3Hrs
Experiment– 5	Demonstrate decision tree algorithm for a regression problem	3Hrs
Experiment– 6	Apply Random Forest algorithm for classification and regression	3Hrs
Experiment– 7	Demonstrate Naïve Bayes Classification algorithm	3Hrs

Experiment– 8	Apply Support Vector algorithm for classification	3Hrs
Experiment– 9	Demonstrate simple linear regression algorithm for a regression problem	3Hrs
Experiment– 10	Apply Logistic regression algorithm for a classification problem	3Hrs
Experiment– 11	Demonstrate Multi-layer Perceptron algorithm for a classification problem	3Hrs
Experiment– 12	How to install and Load CSV files to Python Pandas	3Hrs
Experiment– 13	Implement the K-means algorithm and apply it to the data you selected. Evaluate performance by measuring the sum of the Euclidean distance of each example from its class center. Test the performance of the algorithm as a function of the parameters K	3Hrs
Experiment– 14	Demonstrate the use of Fuzzy C-Means Clustering.	3Hrs
Experiment– 15	Demonstrate the use of Expectation Maximization based clustering algorithm	3Hrs