M.Tech Big Data Analytics

VTR PGE 2023– Curriculum

CBCS - Choice Based Credit System

Department of Computer Science and Engineering
School of Computing

Department of Computer Science and Engineering

VISION

To produce intellectual graduates who could contribute significantly in the analysis, design, development, operation and maintenance of complex software systems for meeting the ever-changing requirements and to compete globally towards professional excellence.

MISSION

The mission of Computer Science and Engineering Department is to

M1: Design curricula for imparting training in adapting newer computing methods and technologies for providing effective and efficient solutions to the existing / new problems.

M2: Inculcate in-depth knowledge of various courses by employing Information and Communication Technology (ICT) based pedagogy methods.

M3: Create a conducive research environment for making technological innovations by the faculty and students.

M4: Provide leadership skills and professional ethics thereby making a prolific career in academics and industry.

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs for CSE)

Graduates of the program will

PEO1: Exhibit the advanced knowledge to apply the analytical and computational approaches to handle big data issues.

PEO2: Demonstrate the skills specific to big data analytics to function productively and professionally.

PEO3: Pursue research activities in related areas involving big data analytics and participate in Lifelong Learning and exhibit proficiency as data analytics professionals.

PEO4: Engage in Critical thinking, Team work, Communication, Leadership along diverse career paths and integrate ethical behavior to become a successful professional in global perspective.

PROGRAM OUTCOMES (POs)

The Graduate will be able:

PO1: Independently carry out research / investigation and development work to solve practical problems.

PO2: Write and present a substantial technical report / document.

PO3: Demonstrate a degree of mastery in the area of Big Data Analytics

PO4: Apply the emerging computing technologies in the field of Data Analytics to conceptualize real world problems and develop appropriate solutions.

PO5: Use advanced statistical tools, specialist software and computing technology effectively to synthesize information for obtaining valid conclusions.

PO6: Demonstrate awareness of ethics and exhibit knowledge of communication, leadership and teamwork to manage projects in multidisciplinary environment and participate in Lifelong learning.

PROGRAM SPECIFIC OUTCOMES (PSO)

Graduates will be able to

PSO1: Design and conduct effective data-driven experiments to meet specific needs within economic, environmental and social constraints

PSO2: Apply the knowledge of data curation to construct data into meaningful structures and Gather Valuable Data Insights to predict the solutions.

COURSE OUTCOMES (COs)

Abilities of the student define in-terms of the course Outcomes (COs) at the end of the course. For all courses in the programmes Course Outcomes (COs) defined as per the Blooms Taxonomy.

M. Tech BIG DATA ANALYTICS
CSBS VTRPGE2023

(With effect from 2023-2024)

Minimum credits required for regular students in various course categories for M.Tech (BDA)

The students shall earn 80 credits in various course categories given below for the award of degree of M.Tech (BDA).

Course Category	Minimum Credits Required – Regular
Program Core	34
Program Elective	18
Open Elective	03

Independent Learning
  i) Technical Writing Tools /Business English
  ii) Research Methodology
	02
	02

Project Work
  i) Phase – I
  ii) Phase – II
  iii) Phase –III
	03
	06
	12

Total	**80**

Vel Tech RangarajanDr.Sagunthala R&D Institute of Science and Technology

Department of Computer Science and Engineering
M. Tech Big Data Analytics - CBCS VTRPGE2023

Program Core (34 Credits)

L – Lecture; T – Tutorial; P – Practical; C – Credits

S.No.	Subject Code	Course Name	L	T	P	C	Pg.No
1	20231CS101	Probability and Statistics	4	0	0	4	8
2	20231CS102	Advanced Data Structures and Algorithms	4	0	0	4	11
3	20231CS103	Advanced Operating Systems	4	0	0	4	14
4	20231CS104	Advanced Database Management Systems	4	0	0	4	17
5	20231CS105	High Performance Computing	4	0	0	4	20
6	20231CS202	Artificial Intelligence	3	0	2	4	23
7	20231CS203	Machine Learning	3	0	2	4	27
8	20231CS204	Big Data Analytics	3	0	2	4	31
9	20231CS301	Advanced Data Structures and Algorithms Laboratory	0	0	2	1	38
10	20231CS302	Advanced Operating Systems Laboratory	0	0	2	1	43
		Total				**34**	

Program Elective (18 Credits)
(within 18 credits 6 Credits from NPTEL as specified by the Department)

L – Lecture; T – Tutorial; P – Practical; C – Credits

S. No	Subject Code	Course Name	L	T	P	C	Pg.No
1	20232CS211	Data Visualization	3	0	2	4	47
2	20232CS201	Deep Learning	3	0	2	4	51
3	20232CS212	Data Science	3	0	2	4	55
4	20232CS203	Computer Vision	3	0	2	4	59
5	20232CS205	Natural Language Processing	3	0	2	4	63
6	20232CS213	Healthcare Analytics	3	0	2	4	66
7	20232CS214	Time series and Forecasting	3	0	2	4	70

8	20232CS215	Image and Video Analytics*	3	0	2	4	
9	20232CS101	Agile Methodologies*	2	0	0	2	
10	20232CS102	Data Privacy and Security	2	0	0	2	74

* The Courses content has yet to be added.

Open Electives (3 Credits)

- ❖ Open electives are the courses offered across the schools to enhance the knowledge breadth and professional competency of the students. The students shall register for appropriate electives offered in other PG programmes based on their area of interest. The courses offered under this category cover the interdisciplinary/ transdisciplinary knowledge.
- ❖ One course (Three credits) of twelve weeks duration shall be taken from the courses offered by the online platform like Swayam, NPTEL.

L-Lecture, T-Tutorial, P-Practical, C-credit

S.No	Course Code	Course Name	L	T	P	C
1	20233CSXXX	Open Elective-1 from NPTEL /Swayam	3	0	0	3

Independent Learning (4 Credits)

Students shall take online platform courses from NPTEL/Swayam/MooC

L-Lecture, T-Tutorial, P-Practical, C-credit

S. No	Course Code	Course Name	L	T	P	C	Pg.No
1	20234CS4XX	i) Technical Writing Tools /Business English	2	0	0	2	
2	20234CS4XX	ii) Research Methodology	2	0	0	2	

Project Work (21 Credits)

S. No	Course Code	Course Name	L	T	P	C	Pg.No
1	20234CS701	Major Project Phase 1	-	-	-	3	
2	20234CS702	Major Project Phase 2	-	-	-	6	
3	20234CS703	Major Project Phase 3	-	-	-	12	

Mandatory

It is mandatory to present/publish at least one paper of the project work in the National/International conference or Scopus Indexed/UGC Care listed Journals.

PROGRAM CORE

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS101	**Probability and Statistics**	**4**	**0**	**0**	**4**

A. Preamble

This course provides in-depth knowledge about probability, sampling, Correlation and Random process.

B. Prerequisite Courses

Nil

C. Course Objectives

Students undergoing this course are expected to

- Familiarize with the basic concepts of probability and random variables.
- Understand the sampling distribution.
- Implement hypothesis testing for small and large samples in real life problem.
- Apply the Correlation and classifications for statistical quality control.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Understand the basic concepts of probability.	K2
CO2	Implement the different types of sampling distribution.	K3
CO3	Apply the concepts of Hypothesis testing and Chi-square testing to solve simple problems.	K3
CO4	Demonstrate the concepts of Correlation.	K3
CO5	Develop the concepts of Random process to solve simple problems.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of COs with Program outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	2			2			2	2
CO2	2			2			2	2
CO3	2			2			2	2
CO4	2			2			2	2
CO5	2			2			2	2

H- Strong; M-Medium; L-Low

F. Course Contents

Unit 1 Introduction   **L-12 Hours**

Axioms of probability - Conditional probability - Total probability - Baye's theorem -Random variable - Probability mass function - Probability density function – Properties -Moments - Moment generating function and their properties - Binomial, Poisson, Geometric, Negative binomial, Uniform, Exponential, Gamma and Weibull distribution and their properties.

Unit 2 Sampling   **L-12 Hours**

Sampling: different types of sampling – Sampling distribution – Sampling distribution of Mean Point Estimation of parameters: general concepts of Estimation – Unbiased estimators – Variance of a point Estimator – Standard error – Method of point estimation (method of moments – method of maximum likelihood) – Statistical intervals for a single sample: confidence interval on the mean of a normal distribution with variance known -Confidence interval on the mean of a normal distribution with variance unknown -Confidence interval on the variance and standard deviation of a normal distribution

Unit 3 Testing of hypothesis   **L-12 Hours**

Hypothesis testing: one sample and two sample tests for means and properties of large samples (z-test), one sample and two sample tests for means of small samples (t-test), F-test for two sample standard deviations – Chi-square test for single sample standard deviation – Chi-square tests for independence of attributes and goodness of fit.

Unit 4 Correlation and Regression   **L-12 Hours**

Correlation – Scatter diagram – Karlpearson coefficient of correlation – calculation of the correlation coefficient for a bivariant frequency distribution – rank correlation – repeated rank – Regression – lines of regression – regression curves – regression coefficients – multiple and partial correlation – coefficient of partial correlation – generalization – multiple correlation.

Unit 5 Random Processes   **L-12 Hours**

Classification – Stationary process – Markov Process – Poisson process – Discrete parameter – Markov chain – Chapman Kolmogorov equations – Limiting distributions.

Total: 60 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Freund, John E.;Johnson, Richard Arnold;Miller, Irwin, Miller & Freund's probability and statistics for engineers, (9th Edn.), Pearson Education, 2017 (Unit 1 to 5).

ii. Online References:

2. Course on "Probability and statistics" Dec 2021[online],Available: https://nptel.ac.in/courses/111/105/111105090.
3. Course on "Probabilities and Statistics course ",Dec 2021[online], Available:https://stanford.edu/~shervine/teaching/cs-229/refresher-probabilities-statistics.

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS102	Advanced Data Structures and Algorithms	4	0	0	4

A. Preamble

This course is designed with the purpose of offering comprehensive insights into the practical implementation and effective utilization of Algorithms and Data Structures. Throughout the course, learners will gain a deep understanding of various algorithms and their real-world applications. They will explore the details of designing, analysing, and implementing algorithms to solve a wide range of computational problems efficiently.

B. Prerequisite Courses
NIL

C. Course Objectives
Learners are exposed to
- Analyse complex problems and design efficient algorithms to solve them.
- Enhance problem-solving skills and algorithmic thinking.
- Equip with techniques to create algorithms optimized in terms of complexity.
- Familiarize with different algorithmic paradigms and to approach diverse problem types using various algorithmic techniques.

D. Course Outcomes
Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Apply the principles of algorithms and advanced data structures to real-world scenarios.	K3
CO2	Solve complex problems using divide and conquer, greedy and dynamic programming techniques.	K3
CO3	Implement appropriate graph-related algorithms for various types of problems.	K3
CO4	Utilize advanced algorithms to effectively address complex computational challenges.	K3
CO5	Make use of P, NP completeness, approximation and randomized principles to provide solutions for optimization problems.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of Cos with Program Outcomes and Programme Specific Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	3		3	3	3	2	3	3
CO2	3		3	3	3	2	3	3
CO3	3		3	3	3	2	3	3
CO4	3		3	3	3	2	3	3
CO5	3		3	3	3	2	3	3

3- High; 2-Medium; 1-Low

F. Course Contents

Unit 1 Introduction to basic data structures and algorithms   L-12 Hours
Role of Algorithms in Computing - Charactering Running Times- Recurrences: Substitution Method - Recursion Tree Method - Master Method - Probabilistic Analysis - Randomized Algorithms. Advanced Data Structures: Binomial Heap - Fibonacci Heap - - Min-Max Heaps – Weight-Balanced Trees -Tries - Splay Trees - B-Trees - Hashing - Disjoint Set - Union Find Operations

Unit 2 Advanced Design and Analysis Techniques   L-12 Hours
Divide and Conquer: Quick and Merge Sorting Techniques - Linear Time Selection Algorithm – Strassen's Algorithm for Matrix Multiplication – Closest Pair and Convex-Hull Problem. Greedy Algorithms: Elements – Activity Selection Problem - Job Sequencing with Deadlines - Optimal Storage on Tapes - Container Loading or Bin Packing Problem - Offline Caching. Dynamic Programming: Elements - Rod Cutting - Matrix-Chain Multiplication - Longest Common Sequence – Flow Shop Scheduling – Coin Changing Problem. Amortized Analysis: Aggregate – Accounting Method - Potential Method.

Unit 3 Graph Algorithms   L-12 Hours
Elementary Graph Algorithms: Breadth First Search – Depth First Search. Minimum Spanning Trees: Algorithms for Kruskal and Prim. Single Source Shortest Path – Bellman Ford – Dijkstra's Algorithm. All Pairs Shortest Path: Floyd-Warshall Algorithm. Network Flow Algorithms: Properties, Ford-Fulkerson Method - Maxflow-Mincut Theorem -Edmonds-Karp Heuristics - Maximum Bipartite Matching.

Unit 4 Advanced Algorithms   L-12 Hours
Number-Theoretic algorithms: GCD algorithm - modular arithmetic - primality testing -Miller Rabin test - Integer factorization - Pollard Rho heuristic. String matching: Rabi-Karp - Knuth-Morris-Pratt algorithms. Probabilistic algorithms: basics of probability theory, pseudorandom generators, Geometric Algorithms: Plane sweep technique, role of sweep- line - status and event-point schedule, line segment intersection problem.

Unit 5 NP-Completeness and Reductions   L-12 Hours

P and NP Problems, NP- completeness and reducibility, NP-completeness proofs - NP-Complete Problems. Approximation Algorithms: vertex cover – traveling salesperson problem – set covering problem - Randomization and linear programming - subset- sum problem.

Total: 60 Hours

G. Learning Resources

i. Text Books:

1. Thomas H.Cormen, Charles E.Leiserson, Ronald L. Rivest and Clifford Stein, "Introduction to Algorithms", Fourth Edition, MIT Press, 2022. [Unit 1,2,3,4]
2. Ellis Horowitz, Sartaj Sahni, Susan Anderson Freed, "Fundamentals of Data Structure in C", Universities Press,2017. [Unit 5]

ii. Reference Books:

1. Anany Levitin, "Introduction to the Design and Analysis of Algorithms", Third Edition, Pearson Education, 2017.
2. V. Aho, J. E. Hopcroft, and J. D. Ullman, "Data Structures and Algorithms", Pearson Education, First Edition, Reprint 2017.
3. Narasimha Karumanchi, "Data Structures and Algorithms Made Easy", Fifth Edition, Career Monk, 2021.

iii. Online References:

1. "Introduction to Data Structures and Algorithms" 31.12.2009. Accessed on Apr. 4. 20, [Online]. Available: https://nptel.ac.in/courses/106/102/106102064/.
2. "Data Structures and algorithms". Accessed on: Apr. 15, 2021 [Online]. Available: https://www.coursera.org/specializations/data-structures-algorithms.
3. "Data structures Visualization" 2011. Accessed on: Apr. 15, 2021 [Online]. Available: https://www.cs.usfca.edu/~galles/visualization/Algorithms.html
4. "Data Structures Fundamentals" 2021, Accessed on: Apr. 20,2021 [Online]. Available: https://www.edx.org/course/data-structures-fundamentals.

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS103	ADVANCED OPERATING SYSTEMS	4	0	0	4

A. Preamble

This course aims to offer students an introductory exploration of advanced concepts in operating systems design. Throughout this course, participants will delve into the architectural aspects and theoretical foundations of OS design, resource allocation, security measures, cluster computing, high-availability considerations, scalability challenges, and various other OS-related topics.

B. Prerequisite Courses

Nil

C. Course Objectives

Learners are exposed to

- Familiarize the concepts of synchronization Mechanisms, Deadlock and Memory Management Techniques.
- Provide an understanding of Distributed Operating System Concepts and Synchronization methods.
- Troubleshoot common issues that may arise when using Windows, such as driver errors, system crashes, and software compatibility issues
- Organize the Files and Disks in UNIX environment

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Understand various synchronization Mechanisms, Process Scheduling and Memory Management Techniques.	K2
CO2	Solve the issues and resolutions related to deadlock detection, centralized, distributed & hierarchical deadlock handling algorithms.	K3
CO3	Explain load distributing algorithms, synchronization techniques , distributed mutual exclusion algorithms & their classifications.	K2
CO4	Apply Diagnostic tools to resolve performance Issues in Windows Operating Systems.	K3
CO5	Implement various File, Disk and Network Management Schemes in Unix Operating System.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of Cos with Program Outcomes and Programme Specific Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	3	3	3	3	3		3	3
CO2	3	3	3	3	3		3	2
CO3	3	3	3	3	3		3	3
CO4	3	3		3	3			
CO5	3	3		3	3			

High- 3; Medium-2; Low-1

F. Course Contents

Unit 1 Fundamentals Of Operating   L-12 Hours

Overview – Synchronization Mechanisms: Critical Section Problem, other Problems, and Language Mechanisms for Synchronization (Monitors, Serializers, Path Expressions) – Processes Deadlocks: Preliminaries, Models of Deadlocks, Models of Resources, Systems with only Reusable Resources

Unit 2 Distributed Deadlock Handling   L-12 Hours

Deadlock handling strategies, Detection: Issues and resolution, Control Organizations, Centralized algorithms (Completely Centralized, The Ho- Ramamoorthy), Distributed algorithms (Path-pushing ,Edge-Chasing, Diffusion computation Based) Hierarchical **algorithms**: The Menasce- Muntz ,Ho-Ramamoorthy Algorithm.

Unit 3 Distributed Scheduling   L-12 Hours

Scheduling: Issues in load distributing, Components of load distributing algorithms, Stability, Load distributing algorithms, Performance Comparison, Selecting load sharing Algorithm-Distributed Mutual Exclusion:Mutual Exclusion algorithms: Lamport's algorithm,The Ricart-Agarwala and Maekawa's Algorithm.

Unit 4 Performance & Storage Management In Windows   L-12   Hours

Windows Architecture-Support, Diagnostic Tools-Monitoring ,Troubleshooting Computer Performance-SysInternal Tools-key stages in troubleshooting-purpose and benefits of the various tools-File Recovery-Application Troubleshooting-methods of file recovery-Troubleshooting Windows Startup, Operating System Service Issues-recovering from startup issues-advanced startup options-failed services- locked account scenarios- Troubleshooting Device Driver Failures, Hardware, Physical Failures. Remediate driver issues, Peripherals. storage technologies and Usage-different storage topologies, FC & iSCSI- RAID redundancy RAID 0, RAID 1, RAID 5, RAID 10-disk types-iSCSI logical units-iSCSI target instances-ATA basic disk, dynamic disk, mount points, file systems, mounting a virtual hard disk, distributed file system.

Unit 5 File,Disk & Network Management In Unix   **L-12   Hours**

Understanding Unix file system permission model, File access criteria - users, groups and permissions. Default permissions, changing file permissions, Importance of Backup, Different Backup tools, Backup and restore services, Basic of DAS, NAS and SAN, Physical disk organization, UNIX partition slices, Mounting and unmounting file systems, Checking and repairing file systems, Monitoring free space, Volume management (LVM) (create, remove, modify, resize), FTP Basic, Configuring FTP. NFS Overview, NFS Installation- Configuration-Troubleshooting. Samba Basics, Features of SAMBA, Installing and Configuring, Linux Squid (Proxy), Features of Sqid, Working of Squid, Configuring and Managing Web Server, Domain Name System, Dynamic Host Configuration Protocol, SSH and NTP, Configure key-based authentication, Synchronize time using other NTP peers, Manage security, Firewall Settings, SELinux, Network and Security services

Total : 60 Hours

G. Learning Resources (in IEEE Format)

   i.   **Text Books:**
   1. MukeshSinghal, NiranjanG.Shivaratri, "Advanced concepts in operating systems: Distributed, Database and multiprocessor operating systems", MC Graw Hill education, 2017.(Unit 1-3)
   2. Thomas, Orin. 'Windows server 2019 inside out". Microsoft Press, 2020 (Unit -4)
   3. Ken Hess, "Practical Linux System Administration" O'Reilly Media, Inc,2023, ISBN: 9781098109035 (Unit-5)
   ii. **Reference Books:**
   1. Abraham Silberschatz, Peter Baer Galvin, Greg Gagne, "Operating System Concepts", Seventh Edition, John Wiley & Sons, 2016.
   2. Andrew S.Tanenbaum, "Modern Operating System", Pearson Edition, Fourth Edition, 2016.
   iii. **Online References:**
   1. Course on "Power of Operating system" ,Aug 2023 [online], Available: https://www.coursera.org/learn/os-power-user
   2. Course on "Distributed Systems" , Aug 2023 [online],Available: https://onlinecourses.nptel.ac.in/noc23_cs72/preview
   3. "Windows Server documentation" , Accessed on Aug, 2023 [Online]. Available: https://learn.microsoft.com/en-us/windows-server/

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS104	**Advanced Database Management Systems**	**4**	**0**	**0**	**4**

A.   Preamble

This course deals with in-depth knowledge about maintain, manage and mitigate database systems in the current real-world scenarios via distributed databases, multi version concurrency control. Summarizing data for data analysis and provide statistic-based decision-making with huge databases like warehouses and big data.

B.   Prerequisite Courses
    Nil

C.   Course Objectives

Learners are exposed to
- Describe the fundamental elements of relational database management systems
- Explain the basic concepts of the distributed and object databases.
- Understand the issues of multi-version granularity locks.
- Know how modern databases formulate SQL queries on data.

D.   Course Outcomes

Upon successful completion of this course, students will be able to

CO Nos.	Course Outcomes	K - Level
CO1	Understand Object-Oriented concepts in the database.	K2
CO2	Implement object-oriented databases using Object Query Language.	K3
CO3	Explain concurrency control with a distributed database.	K2
CO4	Describe the decision-making through the mining process.	K2
CO5	Implement modern data models relating to real-time scenario.	K3

E. Correlation of COs with Program Outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	3		3				3	3
CO2	3		3	3	3		3	3
CO3	3		3	3	3		3	3
CO4	3		3	3	3		3	3
CO5	3		3	3	3		3	3

3- High; 2-Medium; 1-Low

F. Course Contents

Unit 1 Review of Data Model and OODBs   L-12 Hours

Relational model concepts; Relational model constraints and relational database schemas; Update operations, anomalies, dealing with constraint violations, Types, and violations. Enhanced entity relationship modeling, Hierarchical modeling, and Network Modeling. Overview of Object-Oriented Concepts – Objects, Basic Properties. Advantages, examples, Abstract data types, Encapsulation, class hierarchies, polymorphism, examples.

Unit 2 Object and Object-Relational Databases   L-12 Hours

Object Query Language OQL; Conceptual design of Object database. Overview of object-relational features of SQL; Object-relational features of Oracle; Implementation and related issues for extended type systems; syntax and demo examples, Aliasing, Object aliasing. Algebraic operations on databases, Relational Algebra Queries.

Unit 3   Distributed Databases and Concurrency Control   L-12 Hours

Introduction to distributed databases; Distributed DBMS architectures; Storing data in a Distributed DBMS; Distributed catalog management; Distributed Query processing; Updating distributed data; Multi-Version Concurrency Control, Sharding; Distributed transactions; Distributed Concurrency Control and Recovery.

Unit 4   Data Warehousing, Decision Support, and Data Mining   L-12 Hours

Introduction to decision support; OLAP, multidimensional model; Implementation techniques for OLAP; Data Warehousing; Views and Decision Support, View materialization, Maintaining materialized views. Introduction to Data Mining; Counting co-occurrences; Mining for rules; Tree-structured rules; Clustering; Similarity search over sequences; Incremental mining and data streams;

Unit 5   Enhanced Data Models for Some Advanced Applications   L-12 Hours

NoSQL with Object Oriented Concepts- MongoDB-Mongoose; CRUD operation using Mongo, Graph databases- Neo4j- CRUD operation using Neo4j; Neo4j vs Graph; Query modeling using Graph; Introduction to Big Data, Handling Big Data using MongoDB-Map Reduce using MongoDB-Hadoop Node configuration;

Total: 60 hours

H. Learning Resources

i. Text Books

1. Abraham Silberschatz, Henry F. Korth and S. Sudharshan, "Database System Concepts", Seventh Edition, Tata McGraw Hill, 2020. [Units-1,2,3,4,5]
2. Shannon Bradshaw, E Brazil, Kristina Chodorow, "MongoDB: The Definitive Guide - Powerful and Scalable Data Storage", Third Edition, Shroff/O'Reilly Inc., January 2020.[Unit-5].

ii. Reference Books

1. Raghu Ramakrishnan et al, "Database Management Systems", Third Edition, McGraw Hill, 2014.
2. Elmasri Ramez, Navathe S, "Fundamentals of Database System", Seventh Edition, Pearson, 2017.
3. J.D.Ullmann et al, "Database Systems: The Complete Book", Second Edition, Pearson Ed, Inc, 2009.

iii. Online References

1. "Designing local library models", Accessed on: May 05, 2021[online]. Available: https://developer.mozilla.org/en-US/docs/Learn/Server .
2. Michael Hunger, Ryan Boyd & William   Lyon, "The Definitive Guide to Graph Databasesfor the RDBMS Developer", first Ed, E-book, Accessed on: October 27, 2021 [online & Download].   Available: https://neo4j.com/whitepapers/rdbms-developers-graph-databases-ebook/.
3. "Database Management systems", Accessed on: August 29, 2023 [online]. Available [Archive]: https://archive.nptel.ac.in/courses/106/105/106105175/.

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS105	High Performance Computing	4	0	0	4

A. Preamble

The High-Performance Computing (HPC) course is designed to provide students with an in-depth understanding of the principles, methodologies, and technologies employed in the field of high-performance computing. HPC plays a pivotal role in accelerating scientific research, data analysis, simulations, and solving complex engineering problems by harnessing the power of parallel processing and efficient resource management.

B. Prerequisite Course

Advanced Operating Systems

C. Course Objectives

Learners are exposed to:
- Understand the modern computing systems effectively.
- Know Parallel multiprocessors/multicomputer systems.
- Explain HPC systems and their diverse applications.
- Apply CUDA programming model and its applications.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Understand memory technologies and high-performance computing systems.	K2
CO2	Describe parallel computing architectures, programming models, and distributed computing technologies.	K2
CO3	Illustrate High-Performance Computing architecture, components, and applications in scientific and engineering domains.	K2
CO4	Explain the fundamentals of OpenMP parallel programming, including data scoping, work-sharing, synchronization, and reduction techniques.	K3
CO5	Apply the CUDA programming model to solve the real world applications.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of Cos with Program Outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	3			2			2	2
CO2	3			3	2	2	2	2
CO3	3				2	2	2	2
CO4	3		3		2		2	2
CO5	3						2	2

3- High; 2-Medium; 1-Low

F. Course Contents

Unit 1   Multi Core Architecture   L-12   Hours

Memory technologies, hierarchical memory systems, the locality principle and caching, direct-mapped caches, block size, cache conflicts, associative caches, write strategies, advanced optimizations, performance improvement techniques, DRAM – organization, access techniques, scheduling algorithms, and signal systems. Tiled Chip Multicore Processors (TCMP), Network on Chips (NoC), NoC router – architecture, design, routing algorithms, and flow control techniques, Advanced topics in NoC and storage – compression, prefetching, QoS.

Unit 2 Multiprocessor Architecture   L-12 Hours

Classification of parallel computing architectures, Amdahl's Law and Gustafson's Law, SIMD, MIMD, and vector processors Relation to Parallel Multiprocessors/multicomputer Systems, Distributed and Concurrent Programs, Message Passing vs. Shared Memory Systems, Synchronous vs. Asynchronous Executions, Design Issues and Challenges, Distributed Computing Technologies, Clocks and Synchronization, Coordination and Agreement Algorithms, Global State and Distributed Transactions.

Unit 3 Introduction to HPC Architecture   L-12 Hours

Overview of High Performance Computing systems and their applications, Cluster computing vs. supercomputing, GPU computing and accelerators, High-Performance Networking, Performance Optimization and Tuning. Coprocessors, Accelerators in Processor I/O Space, Introduction to GPU, Evolution of GPU Functionality, Modern Graphics Processing Unit Architecture, Heterogeneous System architecture, The Essential OpenACC- Introduction, Programming concepts, Library Calls, Environment Variables, Directives, Mass storage and File systems

Unit 4 OpenMP in HPC   L-12 Hours

Introduction to OpenMP - Parallel execution - Data scoping- OpenMp work sharing for loops-synchronization - Reductions - Loop scheduling - Tasking - OpenMP- parallel Jacobi algorithm-advanced OpenMP: Wavefront parallelization, Efficient OpenMP Programming: Profiling OpenMP Programs, Performance pitfalls, Determining OpenMP overhead for short loops, Serialization, File sharing, Parallel Sparse matrix-vector multiply.

Unit 5 CUDA with HPC   L-12 Hours

CUDA programming model, Basic principles of CUDA programming, CU Concepts of grids, threads, and blocks, Memory handling with CUDA- Shared Memory, Constant Memory, Global Memory, Texture Memory, Multi - CPU and Multi - GPU Solutions – Locality, Algorithms on Multiple GPUs, Single node system, Streams, Multiple – Node Systems Optimization, Libraries, and SDK, Designing GPU based systems.

Total: 60 Hours

G. Learning Resources

i. Text Books
1. T. Sterling, M. Anderson, M. Brodowicz, and W. Gropp, "High-Performance Computing: Modern Systems and Practices," 2020, Morgan Kaufmann [Unit 2,3].
2. Patterson, D.A., and Hennessy, J.L., "Computer Architecture : A Quantitative Approach ", Morgan Kaufmann Publishers, , Inc.2017. [Unit 1,4,5]

ii. Reference Books
1. G. Hager and G. Wellein, "Introduction to High Performance Computing for Scientists and Engineers," 2010, CRC Press.
2. 2. W. E. Nagel, D. H. Kröner, and M. M. Resch, Eds., "High-Performance Computing in Science and Engineering '18," 2019, Springer.

iii. Online References
1. "High Performance Computing".[Online]. Available: "Introduction to Computer Architecture " January, 1, 2017. Accessed on June, 30, 2021 [Online]. Available: https://nptel.ac.in/courses/106/102/106102157/.
2. "GPU Architectures and Programming"April, 17, 2020. Accessed on December,14, 2022Available:https://onlinecourses.nptel.ac.in/noc20_cs41/preview.
3. High-performance Computing [Online] Available :https://www.coursera.org/learn/introduction-high.

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS202	Artificial Intelligence	3	0	2	4

A. Preamble

The primary objective of this course is to introduce the basic principles, techniques, and applications of Artificial Intelligence. In this course, students can explore the various search techniques like basic, advanced and heuristic approach were universal problem solving methods to solve a specific problem and provide the best result. Evolutionary algorithms are used to optimize the AI applications. Goal and constraint-based approach, Knowledge reasoning and inference helps the students to develop the decisions making strategies in intelligent agents and expert systems.

B. Prerequisite Course

20231CS102 – Advanced Data Structures and Algorithms

C. Course Objectives

Learners are exposed to:
- Explore the artificial intelligence techniques by using basic and heuristic search approach.
- Understand the basic principles of artificial intelligence concepts towards problem solving, knowledge representation and learning.
- Explore the artificial intelligence techniques in rules and fact-based approach.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Apply the problem-solving techniques by using State Space and heuristic search approach.	K3
CO2	Use evolutionary and advanced search techniques to provide optimized solution.	K3
CO3	Implement unified planning approach for Constraint-satisfaction problems.	K3
CO4	Apply Logical knowledge representation for rules and fact-based approach.	K3
CO5	Interpret the logic of Uncertainty and Reasoning using resolution methods.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of Cos with Program Outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	3	3		1	1		2	
CO2	3	3	3	1	1		2	
CO3	3	3	3	1	1		2	
CO4	3	3	2	1	2		1	
CO5	3	3	3	1	2		2	

3- High; 2-Medium; 1-Low

F. Course Contents

Unit – 1 Basic and Heuristic Search Techniques   **L-9   Hours**

Artificial Intelligence – Introduction - Evolution of AI- State Space Search: Generate and Test- Simple Search- Depth First Search- Breadth First Search- Comparison of BFS and DFS- Depth Bounded DFS- Depth First Iterative Deepening. Heuristic Search: Heuristic Functions- Best First Search- Hill Climbing- N-Puzzle Problem- Block-world Problem - Variable Neighborhood Descent- Beam Search- Tabu Search- Peak to Peak Methods.
Case Studies – Weather Forecasting.

Unit – 2 Evolutionary and Advanced Search Techniques   **L-9   Hours**

Randomized Search and Emergent Systems: stochastic and evolutionary search algorithms -Simulated Annealing- Emergent Systems- Ant Colony Optimization- Finding Optimal Paths: The Travelling Salesman Problem- Dynamic Programming- Algorithm A*- Iterative Deepening A* - Min-Max- Alpha-Beta Tree search – Monte Carlo Tree Search - Case Studies - Map Navigation - Pruning the OPEN and CLOSED List- Divide and Conquer Beam Stack Search.

Unit – 3 Planning and Strategies   **L-9   Hours**

Planning: planning as search, partial order planning - A Unified Planning Framework - Forward and Backwards State Space Planning- Goal Stack Planning- Plan Space Planning - The STRIPS Domain- Constraint Satisfaction Problem: N-Queens-Cryptarithmetic- Constraint Propagation- Scene Labeling- Higher Order and Directional Consistency- Algorithm Backtracking: Graph Coloring - Case Studies – Robot traversal.

Unit– 4 Knowledge Representation   **L-9   Hours**

Ontologies, foundations of knowledge representation- Logical representation -The Scheme - Frames- Semantic Net- Production rules -Scripts - Inheritance in Taxonomies- Description Logics- Formal Concept Analysis- Conceptual Graphs Knowledge reasoning - Reasoning about objects- relations – events- actions- time and space-reasoning with defaults- Reasoning about knowledge Case studies – Chabot system.

Unit– 5 Knowledge Facets and Logic Inferences L-9 Hours

Knowledge Facets: Intelligent agents: reactive, deliberative, goal-driven, utility-driven, and learning agents- The Wumpus World. Logic and Inferences: Formal Logic- Resolution Method in Propositional Logic- Resolution method in First Order Logic- Deductive Retrieval – Horn Clauses – Forward Chaining- Backward Chaining- Uncertain Knowledge - Probabilistic Reasoning- connection to logic- independence- Bayes rule- Bayesian networks probabilistic inference, Case studies – Multi- Agent Decision Making System.

Total: 45 Hours

G. Laboratory Experiments Total: 30 Hours

Part - 1

Task 1	Implement the Graph search algorithms using Breadth first search and Depth First Search. **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 2	Implement the A * Algorithm to find the optimal path. **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 3	Implement the Mini-Max algorithm that uses recursion to search through the game-tree. **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 4	Write a Program to Implement Tic-Tac-Toe game using Python **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 5	Implement the Ant Colony Optimization to Optimize Ride-Sharing Trip Duration.**Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 6	Solve a Map Coloring problem using constraint satisfaction approach by applying following constraints • Assign each territory a color such that no two adjacent territories have the same color by considering following parameters: Domains, Variables and Constraints. • Apply Basic Greedy Coloring Algorithm: Color first vertex with first color, do following for remaining V-1 vertices. • Consider the currently picked vertex and color it with the lowest numbered color that has not been used on any previously colored vertices adjacent to it. If all previously used colors appear on vertices adjacent to v, assign a new color to it. **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 7	Implement the Monkey Banana Problem in Goal Stack planning using prolog by applying following constraints. Imagine a room containing a monkey, chair and some bananas. That have been hanged from the centre of ceiling. If the monkey is clever enough, he can reach the bananas by placing the chair directly below the bananas and climb on the chair. The problem is to prove the monkey can reach the bananas. The monkey wants it, but cannot jump high enough from the floor. At the window of the room there is a box that the monkey can use. The monkey can perform the Following actions: - 1) Walk on the floor. 2) Climb the box.

3) Push the box around (if it is beside the box).
4) Grasp the banana if it is standing on the box directly under the banana.
Tools- Python, Online Simulator - https://graphonline.ru/en/

Task 8	Implement the N-queen problem using backtracking algorithm. In the 4 Queens problem the object is to place 4 queens on a chessboard in such a way that no queens can capture a piece. This means that no two queens may be placed on the same row, column, or diagonal. **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 9	To Build an Intelligent Chatbot system with Python and Dialog-flow using Interactive Text Mining Framework for Exploration of Semantic Flows in Large Corpus of Text. **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 10	Implement simple fact for following: a. Ram likes mango. b. Seema is a girl. c. Bill likes Cindy. d. Rose is red. e. John owns gold. **Tools- Python, Online Simulator - https://graphonline.ru/en/**
Task 11	Implement the Bayesian networks probabilistic inference. **Tools- Python, Online Simulator - https://graphonline.ru/en/**

Part-2 Use Cases:
Use Case – 1: Implement the Map navigation using Heuristic search approach in python.
Use Case – 2: Develop a simple neural network for simulating logic gate operations.
Use Case – 3: Solve Multiple criteria production scheduling problem using Genetic algorithm.
Use Case – 4: Implement the Text Lemmatization using NLTK Python Package.
Use Case – 5: Implement the Intelligent Chat-Bot system using Python and Dialogflow
Use Case – 6: Implement the Block world problem using STRIPS domain precondition rules in prolog.

<div align="right">

Total: 75 Hours

</div>

H. Learning Resources

i. Text books:
1. Stuart. J. Russell, et al. "Artificial Intelligence by Pearson: A Modern Approach" 4th Ed, 2020 [Unit 1 - 5]

ii. Reference books:
1. E. Rich and K. Knight," Artificial Intelligence", Mc Graw Hill Publishers Inc, 3rd Edition, 2017.
2. Deepak Khemani, "A First Course in Artificial Intelligence", McGraw Hill Education, 2013.

iii. Online Reference Links/ Resources:
1. "Artificial Intelligence :Search Methods for Problem Solving", May. 06, 2014 [Online]. Available: https://nptel.ac.in/courses/106/106/106106126/ .
2. "Artificial Intelligence: Knowledge representation and reasoning" Jan. 19, 2016 [online] Available : https://nptel.ac.in/courses/106/106/106106140/.

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS203	Machine Learning	3	0	2	4

A. Preamble

This course introduces students to new and actively evolving interdisciplinary field of modern data analysis. This course has been designed to introduce the concepts and techniques of machine learning to the students. Students will gain practical mastery over machine learning algorithms and applications of machine learning through a hands-on approach that includes working on business case studies.

B. Prerequisite Course

20231CS104   Advanced Database Management Systems

C. Course Objectives

Students are exposed to
- Apply the concepts of supervised and unsupervised learning algorithms for real time applications.
- Execute the data pre-processing and dimensionality reduction methods for large datasets
- Analyse the dataset for finding the association rules and the frequency of the features for the transactions.
- Interpret and suggest appropriate clustering approaches for various types of problems
- Demonstrate the aspects of simple perceptron and multi-layer perceptron.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Build the appropriate hypothesis for the machine learning problems.	K3
CO2	Design the feature reduced real-world datasets using the dimensionality reduction algorithms	K3
CO3	Apply association rule mining for the Classification and Regression problems.	K3
CO4	Implement various Clustering algorithms for the real time applications	K3
CO5	Solve the neural networks using single and multi-layer perceptron for the real time problems.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of COs with Program outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO 1	PSO2
CO1	3	3	3	3	3	1	1	1
CO2	3	2	3	3	3	3	2	2
CO3	3	3	3	3	3	3	2	2
CO4	3	3	3	3	2	2	2	2
CO5	3	3	3	3	3	3	2	2

3-High; 2-Medium; 1-Low

F. Course Contents

Unit 1 Introduction to Machine Learning　　　　　　　　　　　　　**L-9　Hours**

Machine Learning Applications – Supervised Learning - Learning a Class from Examples – Concept Learning Task – Find-S algorithm – Candidate Elimination algorithm - Vapnik-Chervonenkis Dimension – Probably Approximately Correct Learning – Learning Multiple Classes - Classification – Regression – Parametric Methods – Multivariate Methods

Unit 2　Data Pre-processing　　　　　　　　　　　　`　　　　　　　　　**L-9　Hours**

Data Pre-processing- Needs Pre-processing the Data- Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization - Dimensionality Reduction – Feature Extraction-Variable Selection– Principal Component Analysis - Linear Discriminant Analysis – Quadratic Discriminant Analysis – Factor Analysis –Cross Validation –Resampling methods

Unit 3 Association and Classification　　　　　　　　　　　　　　　**L-9 Hours**

Association Rule Mining- Mining Multilevel Association Rules- Mining Multidimensional Association Rule - Constraint Based Association Mining – Linear Regression - Logistic Regression- Decision Tree Induction - Bayesian Classification – Support Vector Machines –Decision Trees –ID3 - Random Forest – Nearest Neighbor Methods – Univariate Tree – Multivariate Tree – Regression Trees – Bagging and Boosting Classifiers – Adaboost – Xboost – Fine tuning the Classifier parameters.

Unit 4 Clustering Techniques　　　　　　　　　　　　　　　　　　**L-9 Hours**

Cluster Analysis - Types of Clustering methods – Partitioning Clustering- K-means, K- Medoid, K Mode- Density-Based Clustering- Grid based Clustering - Distribution Model Based Clustering - Maximum likelihood Estimation- Expectation - Maximization Algorithm- Hierarchical Clustering-Agglomerative Hierarchical algorithm - Fuzzy Clustering- Fuzzy C-means algorithm - Agglomerative and Divisive Clustering .

Unit 5   Neural Networks   L-9 Hours

Neural Networks – Types of NN – MLP- RecNN- RNN- CNN- LSTM- SNN- Simple Perceptron - Gate implementation- Multi-layer Perceptron – Going Forwards – Going Backwards: Back Propagation Error – Multi-layer Perceptron in Practice – Examples of using the MLP – Overview – Deriving Back-Propagation- Optimization- Hyper parameter tuning – AND -OR- NOT – NOT- XOR gate operations using perceptron

Total: 45 Hours

G. Laboratory Experiments   **Total: 30 Hours**

Part - 1

Task 1	Apply and demonstrate the FIND-S algorithm for finding the most specific Hypothesis based on a given set of training data samples. **Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator.**
Task 2	Implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples. **Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator.**
Task 3	Build a classification model that can effectively analyze and extract features from an image. Apply PCA algorithm to find the appropriate feature. **Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator.**
Task 4	Apply LDA algorithm to select the appropriate data from the given data set. Use XG boost algorithm for classification. **Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator**
Task 5	Write a program to demonstrate the working of the decision tree based on ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample. **Tools: Weka, Python, Scikitlearn, Anaconda navigator.**
Task 6	Write a program to implement the Bagging and boosting model to the real world datasets. Compute the accuracy of the classifier, considering few test data sets. **Tools: Rapid Miner, Python, Scikitlearn, Anaconda navigator.**
Task 7	Apply unsupervised learning to implement K – means clustering to perform cluster analysis, find the optimal number of clusters, identify appropriate features and interpret results. **Tools: Weka, Python, Scikitlearn, Anaconda navigator.**
Task 8	Apply unsupervised learning to implement Hierarchical clustering to perform cluster analysis, find the optimal number of clusters, identify appropriate features and interpret results. **Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator.**
Task 9	Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. **Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator.**

Task 10   Apply back propagation neural network on image data. The idea is to build a Artificial Neural Network model that can effectively analyze and extract features from an image.
Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator

Part-2

Use Cases:
Use case1: Weather classification
Use case2 : Stock market regression analysis
Use case3 : Pet animals clustering application
Use case4 : Biometrics pattern analysis
Use case5 : Logic Gate Implementation using perceptron

Total: 75 Hours

H. Learning Resources

i. Text Books

1. Ethem Alpaydin,, "Introduction to Machine Learning", 4th edition, MIT Press, 2020 [Unit 1- 4]
2. Tom Mitchell, "Machine Learning: A Guide to Current Research", Kluwer academic publisher, 2019. [Unit 5]

ii. Reference books

1. Richard S. Sutton and Andrew G. Barto, "Reinforcement learning: An Introduction", Second Edition, MIT Press, 2019.
2. Oliver Theobald, "Machine Learning for Absolute Beginners", Scatterplot Press, third edition, 2020.
3. Beattie, Guy, "Machine Learning: Start with Machine Learning with all new tips tricks2020.Understand how machine learning can help", 2020.
4. Christopher M Bishob, "Pattern Recognition and Machine Learning", Springer-Verlag New York Inc, 2018

iii. Online References
1. Machine Learning, accessed online April 20, 2021 [online].
   Available: https://www.coursera.org/learn/machine-learning.
2. Machine Learning by Georgia Tech, accessed online April 20, 2021[online],
   Available: https://www.udacity.com/course/machine-learning--ud2.

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS204	Big Data Analytics	3	0	2	4

A. Preamble

This course leverages the insights from big data analytics. It covers various Big Data tools and framework such as Hadoop, Pig, Hive, Spark and GraphX for analytics. It also includes social and Text mining approaches for real time streaming data.

B. Prerequisite Courses

20231CS104 - Advanced Database Management Systems

C. Course Objectives

Learners are exposed to
- Implement the big data through Hadoop and Pig frameworks.
- Apply and visualize structured and unstructured data using data analytics method.
- Interpret and suggest appropriate social and text mining approaches for streaming data.
- Demonstrate on big data tools such as Spark and Hive.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Understand Big Data and its analytics in the real world.	K3
CO2	Apply the Big Data framework like Hadoop & Pig for real world problems.	K3
CO3	Model and visualize the Big Data using analytics and visualization methods.	K3
CO4	Analyze the social streaming data using mining approaches.	K3
CO5	Implement various Big Data Tools such as Spark and Hive for a given problem.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of COs with Program outcomes and Program Specific Outcomes

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO 1	PSO2
CO1	3	3	3	3	3	1	1	1
CO2	3	2	3	3	3	3	2	2
CO3	3	3	3	3	3	3	2	2
CO4	3	3	3	3	2	2	2	2
CO5	3	3	3	3	3	3	2	2

3- High; 2-Medium; 1-Low

F. Course Contents

Unit 1 Introduction to Big Data   L-9 Hours

Big Data- Concepts and Terminology, Big Data Characteristics, Different Types of Data. Big Data Adoption-Business Architecture, Business Process Management Information - Big Data Analytics Lifecycle. Big Data- Online Analytical Processing (OLAP), Extract Transform Load (ETL), Traditional BI, Big Data BI

Unit 2 Hadoop and Pig Framework   L-9 Hours

Hadoop Architecture - Hadoop Distributed File System (HDFS) –YARN – Hadoop I/O – Map Reduce: Developing a map-reduce application – Map-reduce working procedure – Types and Formats - Features of Map reduce: sorting and joins- Pipelining. Introduction to Pig, Parallel processing using Pig, Pig Architecture, Grunt, Pig Data Model-scalar and complex types

Unit 3 Data Modeling and Optimization Problems   L-9 Hours

Exploratory Data Analysis, Data Analytics Methods- Clustering, Association Rules, Regression and Classification, Analytics for Unstructured Data, HBase, Mahout, NoSQL, Cassandra, Data Visualization Techniques- Structured and un structured data visualization

Unit 4 Streaming Data Analysis   L-9 Hours

Text Mining-Events and trends in text streams, Embedding semantics in LDA topic models, Social media analytics-Twitter API-Face book API-Youtube API. Lexicon analysis, Social network analysis. Real-Time Analytics with Spark Streaming and Structured Streaming, introducing real-time processing, Architecture of Spark Streaming, Spark Streaming transformations and actions, Spark Streaming with Kafka and HBase, Advanced concepts of Spark Streaming, Monitoring applications, Structured Streaming.

Unit 5 Big Data Tools   L-9 Hours

Introduction-Hive modules, Data types and file formats, Hive QL-Data Definition and Data Manipulation-Hive QL queries, Hive QL views- reduce query complexity. Hive scripts. Hive QL

Indexes- Aggregate functions, Bucketing vs Partitioning. Overview of Spark, Cluster Management, Application Programming interface (API): Spark Context, Resilient Distributed Datasets, Creating RDD, RDD Operations, and Saving RDD - Lazy Operation – Spark Jobs. Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution.

Total: 45 Hours

G. Laboratory Experiments   Total: 30 Hours

Part- I

Task 1	Configure Hadoop cluster using Docker involves creating multiple containers, each representing a different Hadoop component, and configure them to work together. **Tools: Apache Hadoop, Docker**
Task 2	Perform Data Cube Operations (OLAP Operations) using SQL Queries Rollup Rolldown Slicing Dicing **Tools: Docker, MySQL**
Task 3	Implement the ETL process to generate a refined set of log files in HDFS. Extract relevant information, transform it, and load the results into a new HDFS directory. **Tools: Apache Hadoop, Docker**
Task 4	Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm. To count words in a given file, To view the output file, and To calculate execution time. **Tools: Apache Hadoop, Python,Databricks**
Task 5	Setup Apache PIG and perform Agriculture Dataset Analysis using Docker. Download the dataset containing the Agriculture related data using the below https://www.kaggle.com/abhinand05/crop-production-in-india Executing the PIG queries a. Grouping All Records State wise b. Generate Total Crop wise Production and Area c. Generate Total Crop wise Production and Area d. Average crop production in each district after the year 2000 e. Highest produced crops and details from each State. **Tools: Apache PIG, Docker**
Task 6	Implement Cassandra Shell Commands and CURD operations in CQLSH. **Tools: Cassandra, Docker**
Task 7	Implement Shell Commands using Hbase and Hadoop. **Tools: Apache Hbase & Hadoop, Docker / Data bricks**

Task 8	Implement Stream Processing in Apache kafca.
	Tools: Apache kafca, Java/Python, Docker / Data Bricks

Task 9   Collect any Social Media Data from a Twitter to a Local File with the Topic 'Election2024'. Download and Set Up MongoDB Server and a Client Mongo shell.
Tools: MongoDB, Python: tweepy

Task 10   Retrieve Analytic Information given below from MongoDB created in Task 9:

i. For each "place_type", Find total favorite_count

ii. For each "State_code", find total "retweet_count"

iii. Find out top 10 most frequent topic words of the entire tweet message texts of your collection after lemmatization/stemming and removing all the Stop Words.

Tools: MongoDB, Python: Scipy

Task 11   Implement Hive Commands on e-commerce dataset is stored in a Hive table named ecommerce_data , with the following columns and Perform Aggregation, Filtering data, Sorting, Date functions.

order_id (int)
customer_id (int)
product_id (int)
product_name (string)
quantity (int)
price (double)
purchase_date (string)

Tools: Apache Hive, Docker

Task 12   Implement the following operation on the Hive table, created in the Task 11
   a) **External table:** Create an external Hive table that references the same HDFS location as the original table ecommerce_data.
   b) **Partitioning:** Create a partitioned Hive table on the purchase_date column for better query performance.

Tools: Apache Hive, Docker

   Department of Computer Science and Engineering

Part -II (Use Cases)

UseCase 1: Problem Statement

Design a MongoDB database schema and formulate corresponding queries for a credit card application system. The system should be able to handle user registration, credit card application submission, and retrieval of relevant information. Implement queries to perform tasks such as retrieving user details, processing credit card applications, and managing transaction data efficiently within the MongoDB environment. Ensure the schema supports essential features like user authentication, application status tracking, and transaction history retrieval, adhering to industry security standards and best practices.

Solution Architecture: Design a MongoDB database schema for a credit card application system involves identifying the entities, their relationships, and ensuring that the schema supports the required functionalities efficiently.

UseCase 2: Problem Statement

Develop a scalable e-Commerce application using Databricks and Spark SQL to handle vast amounts of data efficiently. Design and implement Spark RDDs and DataFrames to address critical aspects of the application, such as product catalog management, customer order processing, and inventory tracking. Formulate Spark SQL queries to extract valuable insights, such as top-selling products, customer purchase patterns, and inventory status. Ensure that the system integrates seamlessly with Databricks, utilizing the power of Spark for parallel and distributed processing. Implement optimizations to enhance query performance and provide a robust foundation for advanced analytics in the eCommerce domain.

Solution Architecture: Build Key components include Spark RDDs and DataFrames to represent product catalog, order data, and inventory. For product catalog management, RDDs or DataFrames can be utilized to store essential details like product ID, name, category, and price. Similarly, customer order processing involves organizing order data with details such as order ID, customer ID, product ID, quantity, and order date. Inventory tracking can be handled through RDDs or DataFrames containing product ID and available quantity. Spark SQL queries play a crucial role in extracting insights, including top-selling products, customer purchase patterns, and inventory status.

Use Case 3: Flight Data Analysis using Spark GraphX. To analyze Real-Time Flight data using Spark GraphX, provide near real-time computation results and visualize the results using Google Data Studio.
Use Case – Computations to be done:
1. Compute the total number of flight routes
2. Compute and sort the longest flight routes
3. Display the airport with the highest degree vertex
4. List the most important airports according to PageRank
5. List the routes with the lowest flight costs
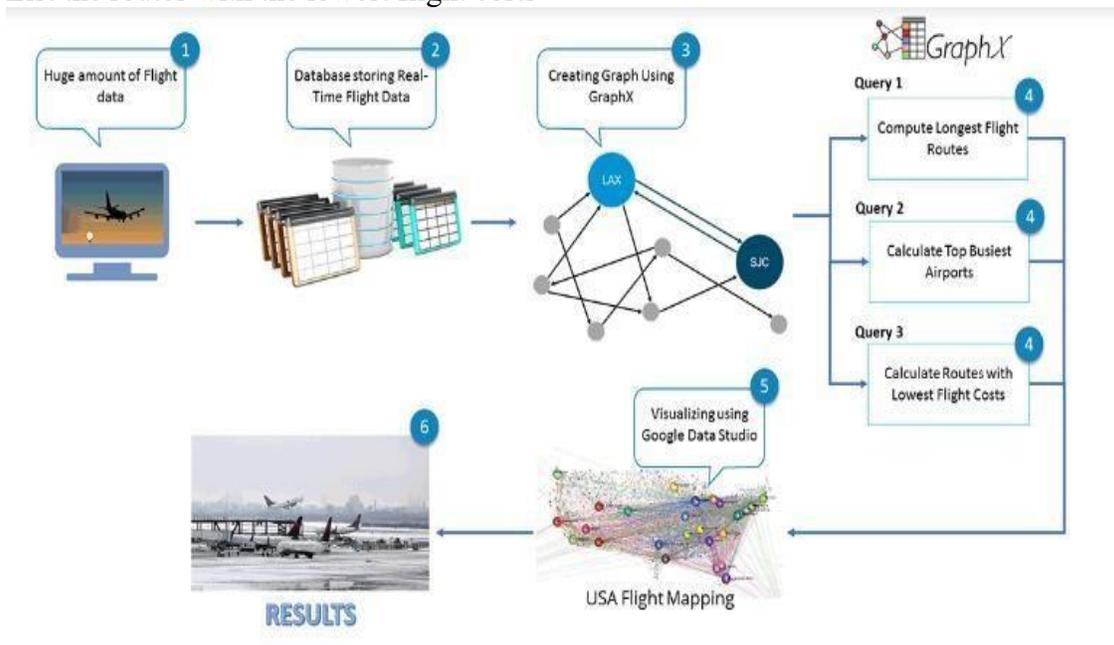
Total: 75 Hours

H. Learning Resources

i.Text Books:

1. Paul Deitel, Harvey M. Deitel, "Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and The Cloud", 1st edition, Published by Pearson (July 14, 2021)
2. Thomas Erl, Wajid Khattak, and Paul Buhler, Big Data Fundamentals: Concepts, Drivers &Techniques, Pearson India Education Service Pvt. Ltd., First Edition, 2016

ii. Reference Books:

1. Tom White, Hadoop: The Definitive Guide, O'Reilly Media, Inc., Fourth Edition, 2015
2. Jason Rutherglen, Dean Wampler, Edward Caprialo, Programming Hive, O'ReillyMedia Inc,2012.
3. Mike Frampton, "Mastering Apache Spark", Packt Publishing, 2015.
4. EMC Education Services, "Data Science and Big Data Analytics Discovering, Analyzing, Visualizing and Presenting Data", Wiley, 2015

iii. Online References:

1. Venkat Ankam "Big Data Analyics" Accessed on August, 26, 2023. [Online]. Available: https://www.packtpub.com/product/big-data-analytics/9781785884696
2. "Big Data Computing" Available from 26-11-2018, Accessed on August, 26, 2023. [Online]. NPTEL Available: https://nptel.ac.in/courses/106104189

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS301	**Advanced Data Structures and Algorithms Laboratory**	**0**	**0**	**2**	**1**

A. Preamble

This course is intended to cultivate the ability to create and evaluate linear and nonlinear data structures. Students gain first hand proficiency in identifying and implementing appropriate data structures for real-world problems. This practical exposure equips them with insights into the tangible applications of these data structures.

B. Prerequisite Courses

Nil

C. Course Objectives

Learners are exposed to:
- Understand the role of algorithms in computing and their significance in solving various computational problems.
- Familiarize with advanced data structures and algorithmic techniques for solving various applications.
- Acquired advanced algorithms to solve specific types of problems

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Implement advanced data structures to efficiently handle complex data storage and retrieval tasks.	K3
CO2	Design efficient algorithmic techniques for computational problems considering complexity factors	K3
CO3	Analyse appropriate graph-related algorithms to solve computing applications	K3
CO4	Make use of number theoretic and probabilistic algorithms to specific types of problems	K3
CO5	Apply randomization and linear programming concepts to address optimization problems with practical solutions	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of Cos with Program Outcomes and Programme Specific Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	3		3	3	3	3	3	2
CO2	3		3	3	3	3	3	2
CO3	3		3	3	3	3	3	2
CO4	3		3	3	3	3	3	2
CO5	3		3	3	3	3	3	2

3- High; 2-Medium; 1-Low

F. Course Contents

Part I:

Task 1: Priority Queue Data Structure

i) Binomial Heap: Implement the decrease-key operation in a binomial heap and analyze its time complexity.

ii) Fibonacci Heap: Implement a mechanism for lazy deletion in Fibonacci heaps.

Task 2: Advanced Data Structures

i) Splay Tree: Implement the operations to find the kth smallest or largest element in a Splay Tree. Test your implementation with various values of k.

ii) B Tree: Implement disk-based operations by tracking the number of page accesses required for various B-Tree operations. Experiment with different node sizes and order values to optimize disk I/O efficiency.

Task 3: Advanced Hashing

i) Implement and analyze cryptographic hash functions like SHA-256.

ii) Implement the Rabin-Karp string searching algorithm that uses hashing to efficiently search for a pattern within a text.

Task 4: Algorithmic Techniques

i) Implement the Strassen's algorithm for matrix multiplication using the divide and conquer approach.

ii) Implement merge sort to count the number of inversions in an array using divide and conquer strategy.

iii) Implement the Huffman coding algorithm for text compression using Greedy technique.

iv) Implement longest common subsequence among multiple strings using dynamic programming.

v) Solve the coin change problem with a twist: instead of finding the number of ways to make change, find the minimum number of coins needed to make a certain amount.

Task 5: Elementary Graph Algorithms

i) Implement Kruskal's algorithm to find minimum spanning tree of a weighted graph.

ii) Implement Prim's algorithm to find minimum spanning tree using priority queue.

iii) Implement the Bellman-Ford algorithm to find the shortest path from a source vertex to all other vertices, even in graphs with negative-weight edges.

Task 6: Network Flow Algorithms

i) Implement the Ford-Fulkerson algorithm to find the maximum flow in a network. Test it on different flow networks and explore different augmenting path strategies (BFS & DFS).

ii) Implement the Edmonds-Karp algorithm, a specific implementation of Ford-Fulkerson using BFS for finding augmenting paths.

iii) Implement max-flow min-cut algorithm to image segmentation problems: to partition an image into different segments while minimizing the cut.

Task 7: Number Theoretic Algorithms

i) Implement an algorithm to generate strong pseudoprimes based on the Miller-Rabin primality test.

ii) Implement Pollard's Rho algorithm for integer factorization

Task 8: Probabilistic & Geometric Algorithms

i) Implement an algorithm generates random passwords using pseudorandom numbers. Users can specify the desired password length and character set

ii) Implement a basic plane sweep algorithm. Use it to solve a simple geometric problem, such as finding intersections among a set of line segments.

Task 9: Randomization and Linear Programming

i) Implement a randomized algorithm to select a random subset of elements from a given array with equal probability

ii) Implement a linear programming model to optimize the construction of a binary search tree with specific access frequencies for the elements. Solve it to find the optimal tree structure.

Part II:

Use Case 1: Emergency Service Dispatch System

An emergency service organization, such as a fire department or medical response team, needs an efficient system to dispatch emergency vehicles to incidents based on their severity and proximity to the incident location. This system must consider both the urgency of the situation and the availability of nearby vehicles. A combination of hashing and priority queues can be employed to create a real-time emergency service dispatch system. Utilize hashing and priority
queue to manage the available emergency vehicles. The priority of each vehicle is determined based on factors such as the vehicle's current location, type of emergency it's equipped to handle, and its availability status.

Develop an algorithm to perform the following operations:
(i) Real-Time Incident Insertion: As new incidents are reported, hash their locations to determine the appropriate grid cell. Simultaneously, insert the incident into the priority queue based on its severity. The more severe incidents will have higher priority in the queue.
(ii) Proximity-based Vehicle Assignment: When an incident is reported, retrieve the nearby grid cells using hashing and identify the vehicles available in those cells. Pop the highest-priority vehicle from the queue and assign it to the incident.
(iii) Dynamic Priority Updates: Update the priority of vehicles in the queue based on their new locations, availability status, and other factors. This ensures that the dispatch system adapts to changing conditions in real-time.

Use Case 2: Network Optimization in Supply Chain Management

A global manufacturing company produces and distributes various products through a complex supply chain network involving multiple suppliers, factories, warehouses, and distribution centers. The company seeks to optimize its supply chain operations to minimize costs and maximize efficiency while meeting customer demands.

Develop and implement an algorithm to perform the following operations:
(i) Optimal Flow Calculation: Apply network flow algorithms such as the Ford-Fulkerson algorithm or the Edmonds-Karp algorithm to calculate the optimal flow of products through the supply chain network. The goal is to satisfy customer demands while minimizing transportation costs.
(ii) Minimum Cost Flow: Incorporate transportation costs as edge weights in the graph. Use algorithms like the Minimum Cost Flow algorithm to find the flow that minimizes the total cost of transporting products across the network.

Use Case 3: Tape Storage Optimization

A media production company creates and stores large volumes of video footage for various projects. The company needs an efficient way to store and retrieve video files from tape storage whileminimizing the number of tapes used and optimizing data retrieval.

Applying a greedy approach to tape storage optimization can help the media production company efficiently manage their video archive. Develop and implement an algorithm to perform the following operations:

(i) Tape Utilization Greedy Strategy:

Implement a greedy strategy that selects the next video file to store based on the remaining space on the current tape. Choose the file that best fits the available space without wasting significant capacity.

(ii) Greedy File Retrieval:

When retrieving files, use a greedy approach that retrieves files from the tapes with the most free space first. This helps prevent prematurely using tapes with significant storage capacity.

Use Case 4: Approximation Algorithm for Traveling Salesman Problem

A delivery company needs to optimize the routes of its delivery trucks to minimize travel time and fuel costs. The company faces the classic Traveling Salesman Problem (TSP), where a salesman needs to visit a set of cities exactly once and return to the starting city. Implement the approximation algorithm in the delivery company's route planning software. This helps the company optimize routes for its delivery trucks in real-world scenarios.

Develop and implement an algorithm to perform the following operations:

(i) Use the Nearest Neighbor algorithm as an approximation technique. Start from a randomly selected city, and in each step, choose the nearest unvisited city to the current city. Continue until all cities are visited, and then return to the starting city.

(ii) To improve the approximation, consider performing multiple iterations of the Nearest Neighbor algorithm with different starting cities and selecting the best solution among them.

Total: 30 Hours

G. Learning Resources

i. Text Books

1. Thomas H.Cormen, Charles E.Leiserson, Ronald L. Rivest and Clifford Stein, "Introduction to Algorithms", Fourth Edition, MIT Press, 2022.
2. Ellis Horowitz, Sartaj Sahni, Susan Anderson Freed, "Fundamentals of Data Structure in C", Universities Press,2017.

ii. Reference Books

1. Anany Levitin, "Introduction to the Design and Analysis of Algorithms", Third Edition, Pearson Education, 2017.
2. V. Aho, J. E. Hopcroft, and J. D. Ullman, "Data Structures and Algorithms", Pearson Education, First Edition, Reprint 2017.
3. Narasimha Karumanchi, "Data Structures and Algorithms Made Easy", Fifth Edition, Career Monk, 2021.

iii. Online References

1. "Introduction to Data Structures and Algorithms". Accessed on Apr. 4. 2020, [Online]. Available: https://nptel.ac.in/courses/106/102/106102064/.
2. "Data Structures and algorithms". Accessed on: Apr. 15, 2021 [Online]. Available: https://www.coursera.org/specializations/data-structures-algorithms.

COURSE CODE	COURSE TITLE	L	T	P	C
20231CS302	**ADVANCED OPERATING SYSTEMS LABORATORY**	**0**	**0**	**2**	**1**

A. Preamble

This course aims to offer students an introductory exploration of advanced concepts in operating systems design. These tasks aim to explore mutual exclusion, inter-process communication, resource allocation, fault tolerance, client-server applications, task scheduling algorithms, and stateful servers. Through these implementations, students will gain hands-on experience in designing and analyzing essential components of modern computing systems and understand their practical implications in real-world scenarios.

B. Prerequisite Courses
NIL

C. Course Objectives
Learners are exposed to
- Understand essential algorithms and concepts related to synchronization Mechanisms, Deadlock avoidance.
- Explore and analyze various aspects of distributed systems and client-server applications
- Use performance and monitor tools to trouble shoot windows server.
- Manage File and Disk in UNIX environment.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Develop algorithms to solve mutual exclusion problem and IPC problems.	**K3**
CO2	Apply Bakers algorithm to ensure the optimal utilization of resources and prevents system from deadlock.	**K3**
CO3	Experiment with Performance and monitor tools to enhance Windows Server Performance.	**K3**
CO4	Organize Files and Disk in UNIX environment.	**K3**
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of Cos with Program Outcomes and Programme Specific Outcomes

COs	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	3	3	3	3	3		3	3
CO2	3	3	3	3	3		3	3
CO3	3	3						3
CO4	3	3						3

High- 3; Medium-2; Low-1

F. Course Contents

TASK 1   Implement the Mutual Exclusion Problem Using Dekker's Algorithm.
Tools: C/C++/ Java

TASK 2   Implement Inter Process Communication Problem (Producer-Consumer / Reader- Writer Problem) Using Semaphores.
Tools: C/C++/ Java

TASK 3   Implement Banker's algorithm.
Tools: C/C++/ Java

TASK 4   Implement and study the incremental/decremented growth of response and service times for different number of client and servers for servicing continuous stream(s) of constant sized messages.
Tools: C/C++/ Java

TASK 5   Implement a name server for registration and identification of services running on another server.
Tools: C/C++/ Java

TASK 6   Implement a client-server application for a computing problem (of exponential complexity).Compare the performance for a local and remote machine of different speeds.
Tools: C/C++/ Java

TASK 7
    a. Performance and Monitor
    b. Monitoring Reliability and Performance- how to use the Performance Monitor and Resource Monitor to monitor system reliability and performance.
    c. Monitoring Events- how to use the Event Viewer tool to monitor events on a Windows Server.
    d. Use monitoring tools to establish a performance baseline- As an IT administrator, you have been tasked with identifying the cause of a slow network performance issue in your Windows Server environment. You need to establish a performance baseline and monitor the network to identify potential bottlenecks and areas for optimization.What monitoring tools and techniques can you use to establish a performance baseline and identify potential network performance issues?
    e. Optimize performance on Windows 10 clients.- In the lab, you have a Windows 10 client device that is running slow and needs to be

optimized for better performance. Use the built-in performance optimization tools and techniques to identify and resolve performance issues, such as disabling unnecessary startup programs, defragmenting the hard drive, and adjusting power settings. Monitor the system performance using Task Manager and other performance monitoring tools to ensure that the changes have a positive impact on the system's overall performance.

TASK 8   Troubleshooting Windows Server
    a. Using Advanced Startup and Windows RE to recover from Boot Failures
    b. Recovering Windows using Reset This PC
    c. Troubleshooting Hardware by Using Windows Memory Diagnostics
    d. Recovering Windows by using a Restore Point

TASK 9   File access criteria - users, groups and permissions

TASK 10
    a. Working with Backup using different Backup Tools
    b. Creating, formating and mounting a partition

G. Learning Resources (in IEEE Format)

i.   Text Books

1. Mukesh Singhal, NiranjanG.Shivaratri, "Advanced concepts in operating systems: Distributed, Database and multiprocessor operating systems", MC Graw Hill education, 2001.
2. Thomas, Orin. 'Windows server 2019 inside out". Microsoft Press, 2020
3. Ken Hess, "Practical Linux System Administration" O'Reilly Media, Inc,2023, ISBN: 9781098109035

ii.   Reference Books

1. Abraham Silberschatz, Peter Baer Galvin, Greg Gagne, "Operating System Concepts", Seventh Edition, John Wiley & Sons, 2016.
2. Andrew S.Tanenbaum, "Modern Operating System", Pearson Edition, Fourth Edition, 2016.1

iii.   Online References

1. Course on "Power of Operating system" ,Aug 2023 [online], Available: https://www.coursera.org/learn/os-power-user
2. Course on "Distributed Systems" , Aug 2023 [online],Available: https://onlinecourses.nptel.ac.in/noc23_cs72/preview
3. "Windows Server documentation" , Accessed on Aug, 2023 [Online]. Available: https://learn.microsoft.com/en-us/windows-server/

PROGRAM ELECTIVE

COURSE CODE	COURSE TITLE	L	T	P	C
20232CS211	Data Visualization	3	0	2	4

A. Preamble

This course explores the design, development, and evaluation of such information visualizations. This course has been meticulously designed to equip students with the essential skills and knowledge needed to master the art and science of visualizing data effectively. Through practical hands-on exercises and real-world applications, students will gain proficiency in creating impactful visualizations

B. Prerequisite Courses

Nil

C. Course Objectives

Learners are exposed to
- Understand the visual perception and core skills for visual analysis.
- Clean and process data for improving EDA analysis.
- Learn to wisely use various visualization structures such as data methods time- varying data, graph and network.
- Create interactive graphs, charts and maps
- Gain best practices for telling stories with data, using interactive databases to create Visualizations.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Utilize data for the wrangling and visualization process.	K3
CO2	Apply Distribution, Correlation, and Multivariate Analysis for the Exploratory data.	K3
CO3	Explore visualization techniques Trees, Graphs, Networks.	K3
CO4	Implement various visualization techniques for the text, document	K3
CO5	Construct a simple information dashboard.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of COs with Program outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6			PSO1	PSO2
CO1	3	2	1	3	2	3			2	3
CO2	2	2	2	1	3	2			3	2
CO3	1	1	1		3	2			2	2
CO4	3	2			3	1			1	
CO5		3			3	2			1	
		H- Strong; M-Medium; L-Low								

F. Course Contents

Unit 1 Introduction to Data Visualization   **L-9 Hours**
Data Types, Data Wrangling, Vision and Color Perception- visually encoding data-Gestalt principles, information overloads- data abstraction- Levels of validation, Concepts and theories of data visualization- visualization pipeline- Visualization process-Visual Variables- Costs and Benefits of visualization.

Unit 2 Distribution, Correlation, and Multivariate Analysis   **L-9 Hours**
Distribution analysis – describing distributions – distribution patterns – distribution displays – distribution analysis best practices – correlation analysis – describing correlations – correlation patterns – correlation displays – correlation analysis techniques and best practices –EDA: multivariate analysis – multivariate patterns – multivariate displays.

Unit 3 Visualizing Data Methods, Time-Series, Network visualization   **L-9 Hours**
Mapping - Time-series analysis, Time-series patterns, Time-series displays - deviation analysis- Displaying Hierarchical Structures, Displaying Arbitrary Graphs/Networks - Issues. - Trees, Hierarchies and Recursion - Networks and Graphs-Acquiring Data-Parsing Data.

Unit 4 Data,Text and Document visualization   **L-9 Hours**
1D,2D,3D,4D visualization, Levels of Text Representations -The Vector Space Model- Single Document Visualizations- Document Collection Visualizations- Extended Text Visualizations- Knowledge Graph- Visualization System: Systems Based on Data Type- Systems Based on Analysis Type- Text Analysis and Visualization- Modern Integrated Visualization Systems.

Unit 5 Information Dashboard Design   **L-9 Hours**
Information dashboard – categorizing dashboards – typical dashboard data – dashboard design issues and best practices– principles of visual perception for dashboard design - Characteristics of dashboards – key goals in visual design process – dashboard display media – designing dashboards for usability – meaningful organization – maintaining consistency – aesthetics of dashboards – testing for usability.

Total : 45 Hours

G. Laboratory Experiments   Total: 30 Hours

Part 1

TASK 1	Exploration of Data Visualization Tools like Tableau, Python libraries based on a preparation of training data samples. **Tools: Google co-lab, Tableau, Anaconda navigator**
TASK 2	To perform and visualize a different perception technique for a single image **Tools: Google co-lab, Anaconda navigator**
TASK 3	Apply the distribution for the data set. **Tools: Google co-lab, Tableau, Anaconda navigator**
TASK 4	Apply the correlation concept for the data set. **Tools: Google co-lab, Tableau, Anaconda navigator**
TASK 5	To visualize and perform Univariate analysis using variables. **Tools: Google co-lab, Python, Tableau, Anaconda navigator.**
TASK 6	To visualize and perform Bivariate analysis using variables. **Tools: Google co-lab, Python, Tableau, Anaconda navigator.**
TASK 7	To visualize and perform Multivariate analysis using Multiple variables involving Multiple measures . Scatterplot Matrix, Parallel Coordinates, Line Graph, Stacked Bar Chart, **Tools: Google co-lab, Python, Tableau, Anaconda navigator.**
TASK 8	To design and perform visualization for Trees • TreeMap, Sun Burst Display **Tools: Tableau, Language :Python**
TASK 9	To design and perform visualization for Graphs and Networks • Force based Layout **Tools: Tableau, Language :Python**
TASK 10	Utilize dataset to implement 2D,3D,4D visualization **Tools: Tableau, Language :Python**
TASK 11	Write a program to generate insight using Text Network Analysis and Visualization **Tools: Wordle, Tag Cloud, WordTree, InfraNodus**
TASK 11	To perform a Document Collection Visualizations. **Tools: Google co-lab, Tableau, Anaconda navigator**

TASK 13	To generate insight dashboard on Text Visualization. **Tools: Google co-lab, Python, Anaconda navigator**
TASK 14	Construct an interactive dashboard for your own **Tools: Google co-lab, Python, Scikitlearn, Anaconda navigator**

Part 2

Use Cases:

Use case1: Weather classification

Use case2 : Energy consumption analysis

Use case3 : Social media insights

Use case5 : Dashboard design of simple project management system

Total: 75 Hours

H.   Learning Resources (in IEEE Format)

i.   Text Books:

1. Matthew Ward, Georges Grinstein, Daniel Keim, "Interactive Data Visualization Foundations, Techniques, and Applications", 2nd Edition, A K Peters/CRC Press, 2021. (Unit 1);
2. Stephen Few, "Now you see it: Simple Visualization techniques for quantitative analysis", Analytics Press, 2009.
3. Stephen Few, "Information dashboard design: The effective visual communication of data", O'Reilly, 2006.

ii.   Reference Books:
1. Scott Murray, "Interactive data visualization for the web", O"Reilly Media, Inc., 2013.
2. Ben Fry, "Visualizing Data:, O"Reilly Media, Inc., 2007.
3. Edward R. Tufte, "The visual display of quantitative information", 2nd edition,GraphicsPress, 2001

iii.   Online References:
1. https://courses.cs.washington.edu/courses/cse512/14wi/
2. https://www.coursera.org/learn/datavisualization

COURSE CODE	COURSE TITLE	L	T	P	C
20232CS201	Deep Learning	3	0	2	4

A. Preamble

In this course, students will navigate the fundamental principles that depend on deep learning, exploring its architecture, algorithms, and real-world applications. By laying a strong foundation and covering the basics of neural networks, activation functions, and optimization algorithms. This course then progresses to more advanced topics, including convolutional neural networks (CNNs) for image analysis, recurrent neural networks (RNNs) for sequential data, and transformer-based models for natural language processing.

B. Prerequisite Courses

20231CS203 - Machine Learning

C. Course Objectives

Learners are exposed to
- Make use of the basic concept of deep learning.
- Choose the efficient optimization model for improving the model.
- Build the CNN model for an application.
- Construct the Convnet and Transfer learning model.
- Apply the RNN model for time series forecastin.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Make use of basic concepts of deep learning and fundamentals of neural network.	K3
CO2	Choose the suitable optimization technique to improve the model	K3
CO3	Build the CNN architecture to solve real world problems, applying knowledge of layer configurations, filter sizes and strides	K3
CO4	Implement an appropriate Convnet model for a given application and apply transfer learning techniques to fine-tune pre-trained CNN models for specific image recognition tasks	K3
CO5	Apply RNN architectures to sequence modeling tasks, selecting appropriate hyperparameters, and justifying design choices for tasks	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E.   Correlation of COs with Program outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	2	2	3		2	1	2	
CO2	2	3	3		2	2	2	
CO3	3	3	3		3	2	3	2
CO4	3	3	3		3	2	3	3
CO5	3	3	3		3	3	3	3
H- Strong; M-Medium; L-Low								

F.   Course Contents

Unit 1 Introduction to Deep Learning   **L-9 Hours**

Concept of Deep Learning (DL) - Deep Learning Parameters - Deep Learning Frameworks - Introduction to Neural Networks: Feed-forward Networks, Deep Feed-forward Networks – Multilayer Perceptron, Back-propagation - Regularization for Deep Learning - Applications

Unit 2 Optimization Techniques   **L-9 Hours**

Introduction - Optimization in DL - Parameters - Gradient Descent (GD): Variants of GD, Momentum - Types: Stochastic GD - RMSProp - AdaDelta - AdaGrad - Adam Optimizers - Generalization in Neural Networks.

Unit 3 Deep Learning Architectures   **L-9 Hours**

Deep Learning Architectures - Convolutional Layer - Kernel – Padding - Activation Function - Pooling Layer - Basic Terminologies - Types of errors, bias-variance tradeoff, overfitting and underfitting - Normalization and Data Augmentation - Applications - Issues and Challenges of Deep Learning.

Unit 4 Deep Convolutional Neural Network   **L-9 Hours**

Introduction - Convolution Neural Network (CNN) Architecture - CNN representations: invertibility, stability, invariance, covariance/invariance: capsules and related models - Types of Convnet - AlexNet, VGG, ResNet - Weight Initialization - Xavier / Golort - Auto encoders - Transfer Learning: Generative Adversarial Network.

Unit 5 Recurrent Neural Network   **L-9 Hours**

Introduction - Recurrent Neural Network (RNN) Architecture - Terminologies: Backpropagation Through Time, Types of RNN - Variants of RNN - Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional LSTM - Generative models - Restrictive Boltzmann Machines (RBMs) - Representation Learning: Regularization & Autoencoder, Word Representation, Using Word Embeddings, Properties of Word Embeddings.

Total : 45 Hours

G.   Laboratory Experiments   **Total: 30 Hours**

<div align="center">

Part 1

</div>

TASK 1	Implement of binary image classification using Convolutional Neural Network using tensorflow or pytorch **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 2	Implement an optimization technique using tensorflow or pytorch **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 3	Implement a convolutional neural network (CNN) using a framework using TensorFlow or PyTorch to classify images from a popular dataset like CIFAR-10 or MNIST. **(Tools: Google Colab, Python, Anaconda Navigator**
TASK 4	Implement the AlexNet architecture and train it for image classification on the CIFAR-10 dataset using tensorflow or pytorch. **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 5	Implement of Recurrent Neural Network (RNN) for text classification using a framework using TensorFlow or PyTorch **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 6	Build a CNN model to classify a specific dataset while incorporating data augmentation techniques using tensorflow or pytorch. **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 7	Implement the YOLO (You Only Look Once) architecture to create an object detection model using tensorflow or pytorch. **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 8	Implement a basic GAN to generate images of a particular type, such as faces, using a dataset like CelebA (tensorflow or pytorch) **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 9	Time Series Prediction with LSTM using tensorflow or pytorch. **(Tools: Google Colab, Python, Anaconda Navigator)**
TASK 10	Build an autoencoder-based anomaly detection system using tensorflow or pytorch. **(Tools: Google Colab, Python, Anaconda Navigator)**

<div align="center">

Part 2

</div>

Use Cases:

Use Case 1: Dog vs Cat image classification - Pet adoption shelters receive a large number of animals, including dogs and cats, on a daily basis. Efficiently categorizing these animals is essential for various purposes, such as managing inventory, matching pets with potential adopters, and ensuring proper care and attention.

Use Case 2: Time series forecasting - A retail store, XYZ Supermart, sells a wide range of products, including groceries, electronics, and clothing. The store experiences fluctuations in demand for various items due to factors like seasonal trends, promotions, and changing customer preferences.

                                                          Total: 75 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, " Deep Learning", published by MIT Press [Online]. [Unit 1-5]
2. Charu C. Aggarwal, "Neural Networks and Deep Learning: A Textbook", Springer; 1st ed. 2018 edition [Online]. [Unit 1,3,5]

ii. Reference Books:

1. Kelleher, John D.. Deep Learning. United Kingdom: MIT Press, 2019. [Unit 3,4]
2. Nikhil Buduma.,"Fundamentals of Deep Learning", O'Reilly Media, 2017. [Unit-2,3,5]

iii. Online References:

1. "An introduction to deep learning", Accessed on August 14, 2023, Published on March 2, 2020, [Online], Available https://developer.ibm.com/learningpaths/get-started-with-deep-learning/an-introduction-to-deep-learning/

2. "Deep Learning - IIT Ropar", Accessed on August 14, 2023[Online], Available https://nptel.ac.in/courses/106106184

COURSE CODE	COURSE TITLE	L	T	P	C
20232CS212	**Data Science**	**3**	**0**	**2**	**4**

A.   Preamble

Data science is a multidisciplinary field that uses scientific methods, processes, algorithms, and systems to extract insights and knowledge from structured and unstructured data. It combines expertise from various domains such as statistics, mathematics, computer science, and domain-specific knowledge to analyze and interpret complex data sets.

B. Prerequisite Course

20231CS102 - Probability and Statistics

C. Course Objectives

Learners are exposed to:
- Identify general statistical techniques for data analysis
- Determine how to summarize data to present information using statistics.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Acquire and clean diverse datasets, and execute exploratory data analysis to visualize the data.	K3
CO2	Apply probability and statistics, implement regression analysis, and interpret results for informed decision-making.	K3
CO3	Implement supervised learning algorithms to train and evaluate the model.	K3
CO4	Interpret feature engineering, optimize models, and apply time series analysis for forecasting.	K3
CO5	Use data science for decision-making, culminating in a comprehensive capstone project considering emerging trends and ethical considerations.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of Cos with Program Outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2
CO1	3	2	2		3				2			3	2
CO2	3	2	2		3				2			3	2
CO3	3	2	2		2				2			2	2
CO4	2	2	2		3				2				2
CO5	2	2	2		2				2				2

3- High; 2-Medium; 1-Low

F. Course Contents

UNIT – I Introduction to Data Science   **L-9**

Overview of Data Science: Definition and scope of data Science-Historical perspective and Evolution-Applications in various industries- Data Acquisition and Cleaning: Data sources and Types-Data Collection Methods-Data cleaning and preprocessing techniques-Exploratory data Analysis: Descriptive statistics- Data visualization techniques-EDA tools and practices.

UNIT – II Statistical Foundations for Data Science   **L-9**

Probability and Statistics: Probability Distributions-Statistical Inference-Hypothesis testing- Regression Analysis: Simple and multiple Regression-Model evaluation and Interpretation-Practical applications- Apply regression analysis for modeling relationships in data.

UNIT – III Machine Learning Basics   **L-9**

Introduction to Machine Learning: Types of machine learning (supervised, unsupervised, reinforcement learning)-Model training and evaluation-Supervised Learning Algorithms: Linear regression, decision trees, and random forests-Support Vector Machines (SVM)-Neural networks and deep learning basics

UNIT – IV Feature Engineering and Model Optimization   **L-9**

Feature selection and extraction-Hyper parameter tuning-Cross-validation techniques- Time Series Analysis-Time series data characteristics-Forecasting methods-Seasonality and trend analysis

UNIT – V Data Science for Decision Making   **L-9**

Decision Support Systems (DSS):Introduction to decision support systems-Components and architecture of DSS-Role of data science in enhancing decision support systems -Business Intelligence and Analytics: Overview of business intelligence (BI)-Key performance indicators (KPIs) and metrics-Data visualization for business insights

Total: 45 Hours

G. Laboratory Experiments   **Total: 30 Hours**

Part - 1

Task 1 Implement a Python program for acquiring and cleaning a dataset using the pandas library.

Task 2. Perform exploratory data analysis using descriptive statistics and visualization techniques

Task 3 Use Python to calculate probability and statistical analysis.

Task 4 Use Python to apply regression analysis techniques and evaluate models.

Task 5 Develop a model to gain hands-on experience in understanding types of machine learning and the basic process of model training.

Task 6 Implement three different classification algorithms, such as Support Vector Machines (SVM), Decision Trees, and Random Forest.Train each algorithm on 70% of the dataset and evaluate their performance on the remaining 30%. Use metrics such as accuracy, precision, recall, and F1 score for evaluatio

Task 7 Explore the impact of hyperparameter tuning on the performance of each algorithm and discuss any observed improvements or trade-offs.

Task 8 Initiate the exploration of the historical sales data, utilizing visualizations to uncover patterns, trends, and potential seasonality embedded in the dataset.

Task 9 Design and implement a basic decision support system using Python to assist in optimizing investment strategies.

Task 10 Apply analytics techniques using NumPy and SciPy to derive meaningful insights from the sales data. How can statistical analysis help identify sales patterns, correlations, and outliers within the dataset?

Tools- Jupyter Notebook, Google Colab

Libraries – NumPy, SciPy, Pandas, Matplotlib, Seaborn

Part-2 Use Cases:

Use Case – 1: A regional healthcare provider aims to improve patient care and resource allocation. The decision support system is designed to assist healthcare professionals in making informed decisions..

Use Case – 2: A retail company wants to boost sales and enhance its overall business strategy. The BI system is used to analyze sales performance and identify areas for improvement.

Use Case – 3: A financial institution aims to enhance its fraud detection capabilities. The DSS is designed to analyze transaction data and identify potential fraudulent activities.

Use Case – 4: A manufacturing company seeks to optimize its production processes. The DSS is designed to analyze data from the production line and provide insights for process improvement.

Use Case – 5: A telecommunications company wants to reduce customer churn and enhance customer satisfaction. The BI system is used to analyze customer data and develop strategies for retention.

Total:75 Hours

G. Learning Resources

  i.   **Text books:**
   1. David Cielen, Arno D. B. Meysman, and Mohamed Ali, "Introducing Data Science", Manning Publications, 2016.

2. Data Science for Business by Foster Provost, Tom Fawcett Released August 2013, Publisher(s): O'Reilly Media, Inc.ISBN: 9781449361327

ii.   Reference books:

1. Alice Zheng and Amanda Casari. 2018. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists (1st. ed.). O'Reilly Media, Inc.
2. Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer, 2017.

iii.   Online Reference Links/ Resources:

1. [Data Science for Engineers - Course (nptel.ac.in)](#)

2. [Applied Data Science with Python Specialization [5 courses] (UMich) | Coursera](#)

COURSE CODE	COURSE TITLE	L	T	P	C
20232CS203	Computer Vision	3	0	2	4

A. Preamble

This course offers practical insights into computer vision. Equipping students with algorithms, methodologies, and concepts, it empowers the creation of potent visual systems. From image processing to 3D vision and motion analysis, master object recognition, segmentation, and motion estimation, gaining skills to tackle real-world challenges.

B. Prerequisite Courses

20231CS203 - Machine Learning

C. Course Objectives

Learners are exposed to
- Understand image formation models and geometric image processing comprehensively.
- Illustrate the basic techniques of image filtering and edge detection concepts.
- Develop expertise in image segmentation and classification methods.
- Implement a variety of techniques to explore grouping and model fitting.
- Explore 3D representations, shape recovery, and motion estimation principles in depth.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Understand of image formation models, camera calibration, and color representations to analyze and interpret images accurately.	K2
CO2	Solve image enhancement and feature extraction challenges through the application of filters and edge detection algorithms.	K3
CO3	Demonstrate the ability to segment images into meaningful regions and classify objects.	K3
CO4	Apply robust model fitting techniques to interpret data accurately.	K3
CO5	Implement acquired skills proficiently to estimate three-dimensional shape information from images and motion patterns.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. **Correlation of COs with Program outcomes and Programme Specific Outcomes:**

Cos	PO1	PO2	PO3	PO4	PO5		PO10	PO11	PSO1	PSO2
CO1	3	1	1	1	1		1	2	2	1
CO2	3	3	3	2	3		1	2	3	2
CO3	3	3	2	2	3		1	2	3	2
CO4	2	3	3	2	3		1	2	3	2
CO5	2	3	3	2	3		1	2	3	2
H- Strong; M-Medium; L-Low										

F. **Course Contents**

Unit 1 Introduction **L-9** **Hours**
Introduction, Image Formation Models: Geometric image formation, Pinhole Camera Model, Camera Calibration Techniques: Intrinsic and Extrinsic Parameters, Binocular imaging systems, Photometric Stereo, Shape from shading, Color Representations, Color Spaces, Inference from Color.

Unit 2 Image Filtering and Edge Detection **L-9** **Hours**
Introduction to Image Filtering, Spatial Domain Filtering, Frequency Domain Filtering, Non-linear Filters, Sampling and Aliasing, Edge Detection Concepts, Gradient-Based Edge Detection, Canny Edge Detection Algorithm, Edge Linking and Edge Hysteresis, Boundary Detection Techniques.

Unit 3 Image Segmentation and Classification **L-9** **Hours**
Human Vision: Grouping and Gestalt, Applications: Background Subtraction, Shot Boundary Detection, Interactive Segmentation, Forming Image Regions, Watershed Algorithm, Segmentation Using K-means, Classifying Images: Encoding Layout with GIST Features, Summarizing Images with Visual Words, Spatial Pyramid Kernel, Classifying images of single objects.

Unit 4 Grouping and Model Fitting **L-9** **Hours**
Hough Transform, Fitting Lines and Planes, Fitting Curved Structures, Robustness - M-Estimators, RANSAC: Searching for Good Points, Fitting Using Probabilistic Models, Motion Segmentation by Parameter Estimation - Optical Flow and Motion, Flow Models, Motion Segmentation with Layers.

Unit 5 3D Representations and Motion Estimation **L-9** **Hours**
3D Representations: Introduction, 3-D Vision, Projection Schemes, Shape from Shading, Photometric Stereo, Three-Dimensional Object Recognition Schemes, Horaud's Junction Orientation Technique. Motion estimation: Translational alignment, parametric motion, Optical flow, Layered motion.

Total : 45 Hours

G. Laboratory Experiments   Total: 30 Hours

Part 1

TASK 1	Understand the pinhole camera model and camera calibration techniques (intrinsic and extrinsic parameters). **(Tools: Python)**
TASK 2	Implement linear and non-linear camera calibration algorithms using python language. **(Tools: Python)**
TASK 3	Explore various color representations and color spaces in images. **(Tools: Python)**
TASK 4	Apply spatial domain filters (mean, median, Gaussian) to filter images and remove noise. **(Tools: Python)**
TASK 5	Perform frequency domain filtering using Fourier Transform. **(Tools: Python)**
TASK 6	Detect the edges of the given image using gradient-based approach. **(Tools: Python)**
TASK 7	Use basic clustering methods (K-means) and the watershed algorithm for image segmentation. **(Tools: Python/ MATLAB)**
TASK 8	Encode image layouts using GIST features and summarize images with visual words **(Tools: Python/ MATLAB)**
TASK 9	Implement spatial pyramid kernel for image classification. **(Tools: Python/ MATLAB)**
TASK 10	Implement Hough transform to detect lines and circles in images. **(Tools: Python/ MATLAB)**
TASK 11	Fit lines and planes to 2D and 3D data using linear regression techniques. **(Tools: Python/ MATLAB)**
TASK 11	Fit curved structures using non-linear model fitting with M-estimators and RANSAC. **(Tools: Python/ MATLAB)**
TASK 13	Use 3D representations and shape-from-shading to recognize three-dimensional objects. **(Tools: Python/ MATLAB)**
TASK 14	Implement Horaud's junction orientation technique for object recognition. **(Tools: Python/ MATLAB)**
TASK 15	Apply computer vision techniques for digital heritage, facial animation, 3D model capture, photo tourism, 3D photography, animating pictures, and video-based walkthroughs. **(Tools: Python/ MATLAB)**

Part 2

Use Cases:

Use Case 1: Human Iris Location for determining the gaze direction
Use Case 2: Real-time Facial Expression Recognition for Human-Computer Interaction
Use Case 3: Lane Detection and Departure Warning System
Use Case 4: Application of Computer Vision in Visual Hulls: Tracing Intersection Curves, Clipping Intersection Curves, Triangulating Cone Strips, Carved Visual Hulls
Use Case 5: 3D Object Reconstruction for Cultural Heritage Preservation

Total: 75 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Forsyth and Ponce, "Computer Vision – A Modern Approach", Second Edition, Prentice Hall, 2011. (UNIT I - IV)
2. Richard Szeliski, "Computer Vision: Algorithms and Applications", Second Edition,Springer, 2022. (UNIT V)

ii. Reference Books:

1. E. R. Davies, ―Computer & Machine Vision‖, Fourth Edition, Academic Press, 2012.
2. Richard Hartley and Andrew Zisserman, "Multiple View Geometry in Computer Vision", Second Edition, Cambridge University Press, 2004.

iii. Online References:

1. "Computer Vision Basics" on OpenCV Tutorials. Apr 2, 2021. Accessed on: Aug. 21, 2023 [Online].
   Available: https://docs.opencv.org/4.5.2/d9/df8/tutorial_root.html
2. "Introduction to Computer Vision", Accessed on: Aug. 21, 2023 [Online].
   Available:https://www.udacity.com/course/introduction-to-computer-vision--ud810
3. "Computer Vision" on Stanford University's Open Classroom. Accessed on: Aug. 21, 2023 [Online].
   Available:http://openclassroom.stanford.edu/MainFolder/CoursePage.php?course=MachineLearning

COURSECODE	COURSE TITLE	L	T	P	C
20232CS205	**Natural Language Processing**	3	0	2	4

A. Preamble

Natural language processing deals with written text. Students will learn how to process written text from basic of fundamental knowledge starts with Finite automata, Regular expression and probabilistic model with n-grams. This course also covers basis of semantic analysis and discourse analysis and drives it to machine translation

B. Prerequisite Courses

20231CS102   Advanced Data Structures and Algorithms

C. Course Objectives (verbs)

Student will be able to
- Provide the student with knowledge of various levels of analysis involved in NLP
- Understand the applications of NLP
- Gain knowledge in automated Natural Language Generation and Machine Translation
- Understand Deep learning concepts for NLP.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

COs	Course Outcomes	K-Level
CO1	Comprehend the fundamental concepts of Natural Language Processing.	K3
CO2	Apply the concept of Probabilistic model of defining language and POS tagging for NLP applications	K3
CO3	Develop the Syntactic parser and Semantic Parser for Translation for word forms	K3
CO4	Interpret the concept of text analysis, summarization and extractions for Information Extraction.	K3
CO5	Apply the concepts of machine translation and deep learning for NLTK Modelling.	K3
Knowledge Level (Based on revised Bloom's Taxonomy)		
K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E.   Correlation of COs with Program outcomes and Programme Specific Outcomes:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	2		2	2			2	
CO2	2		2	2			2	2
CO3	2		2	2			2	2
CO4	2		2	2			2	2
CO5	2		2	2			2	2

H- Strong; M-Medium; L-Low

F. Course Contents

Unit 1 Introduction to Natural Language Processing   L-9 Hours

History of NLP, Components of NLP, Applications of NLP, Phases of NLP, NLP APIs, NLPLibraries, Knowledge in Speech and Language processing, Ambiguity and models and algorithm,language and understanding,brief history. Regular Expressions, Extracting Terms fromTokens, Vector Space Representation and Normalization, Similarity Computation in Text

Unit 2 Language Modelling and Part-of-Speech (POS) Tagging   L-9 Hours

Unigram Language Model, , Counting Words in Corpora, Simple (Unsmoothed) N-grams, Smoothing, Back off, Deleted Interpolation, N-grams for Spelling and Pronunciation, Entropy Natural Language Generation, Parts of Speech Tagging, Morphology, Named Entity Recognition, Overview of Hidden Markov Models.

Unit 3 Words and Word Forms   L-9 Hours

Context-Free Grammars for English, Lexicalized and Probabilistic Parsing, Semantic Analysis,Lexical Semantics, WordNet: A Database of Lexical Relations, Bag of words, skip-gram,Continuous Bag-Of-Words, Embedding representations for words Lexical Semantics, WordSense Disambiguation, Knowledge Based and Supervised Word Sense Disambiguation

Unit 4 Text Analysis, Summarization and Extractions   L-9 Hours

Sentiment Mining, Entity Linking, Text Classification, LDA & Factorization - Text Summarization, Information Extraction, Named Entity Recognition, Relation Extraction, Question Answering in Multilingual Setting; NLP in Information Retrieval, Vector space model,Cross-Lingual IR

Unit 5 Machine Translation and Deep Learning   L-9 Hours

Need of MT, Problems of Machine Translation, MT Approaches, Statistical Machine Translation(SMT), Parameter learning in SMT (IBM models) using EM, Encoder-decoder architecture,Neural Machine Translation Statistical Language Models, Kernel Methods, Word-ContextMatrix Factorization Models, Neural Language Models, Recurrent Neural Networks, SentimentAnalysis, Opinion Mining.

Total: 45 Hours

Lab Task: 30 Hours

Part - I

1. Introduction to Python NLTK. **spaCy,**Libraries (CO1)
2. Performing Word Analysis – NLTK &spaCy (CO1)
3. Analyze algorithms for Word Generation - NLTK &spaCy(CO2)
4. Morphology is important factor for word Embedding. Develop algorithms for findingmorphology of Word Documents. NLTK &spaCy, Gensim (CO3)
5. Develop N-Grams for Word Document - NLTK &spaCy, Gensim (CO3)
6. Perform N-Grams Smoothing in Word Documents - NLTK &spaCy, Gensim (CO3)
7. Explore the POS Tagging using Hidden Markov Model - NLTK &spaCy, Gensim (CO3)
8. Use POS Tagging using Viterbi Decoding - NLTK (CO4)
9. Building POS Tagger for unstructured Web Documents [PyTorch, Keras] (CO4)
10. Chunking for Web Documents [PyTorch, Keras] (CO5)
11. Building Chunker for Web Documents [PyTorch, Keras] (CO5)
12. Python Visualizations Libraries – matplotlib, seaborne (CO5)

Part – II [CO1 – CO5]

13. Game Winning Prediction (Cricket)
14. Machine Translation from English-Hindi.
15. Query Expansion for Information Retrieval.
16. Emotion detection for texts.
17. Semantic Search Engine for Indian Languages.
18. Large Hindi Corpora
19. Crisp Query relevant Summary of each retrieved webpage
20. Semantic Query-Webpage Relevance

Total Hours: 75 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Daniel Jurafsky, James H. Martin, ‒Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech‖,, Pearson Publication, 2014.(All 5 Units)
2. Steven Bird, Ewan Klein and Edward Loper, ―Natural Language Processing with Python‖, First Edition, O_Reilly Media, 2009.

ii. Reference Books:

1. Breck Baldwin, ‒Language Processing with Java and LingPipe Cookbook‖, Atlantic Publisher, 2015
2. Charu C.Aggarwal, ―Machine Learning for Text‖-bySpringer,2018 edition
3. Nitin Indurkhya and Fred J. Damerau, ―Handbook of Natural Language Processing,Second Edition, Chapman and Hall/CRC Press, 2010.

iii. Online References:

1. ‒Natural Language Processing‖, Mar. 2015. Accessed on: Apr. 16, 2021 [Online]: https://onlinecourses.nptel.ac.in/noc19_cs56
2. ‒Natural Language Processing‖ Jun 2016. Accessed on: Apr. 16, 2021 [Online]. https://www.udacity.com/course/natural-language-processingnanodegree
3. ‒Natural Language Processing‖ Jul. 2, 2018. Accessed on: Apr. 16, 2021 [Online]. https://www.coursera.org/learn/language-processing.

COURSE CODE	COURSE TITLE	L	T	P	C
20232CS213	Healthcare Analytics	3	0	2	4

A. Preamble

This course explores the concept of clinical intelligence and the role of analytics in supporting a data-driven healthcare system. The aim of this course is to focus beyond data collection, to analyze available data and making it into actionable information.

B. Prerequisite Courses

Nil

C. Course Objectives

Learners are exposed to
- Explore the various forms of electronic health care information.
- Learn the techniques adopted to analyze health care data.
- Understand the predictive models for clinical data.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Understand the fundamental concepts in Healthcare Analytics	K2
CO2	Apply various segmentation and feature extraction to perform biomedical image analysis.	K3
CO3	Analyze the various types of biomedical signal analysis techniques.	K3
CO4	Perform data analysis using natural language processing and data mining.	K3
CO5	Make use of visualization techniques to visualize medical data	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of COs with Program outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	2	2						
CO2	3	2		2			2	2
CO3	3	2			2		2	3
CO4	2	2			2			
CO5	2	3			3		2	3
H- Strong; M-Medium; L-Low								

F. Course Contents

Unit-1 Healthcare, and Data Analytics   L-9 Hours

Introduction – Healthcare Data Sources and Basic Analytics –Resources for Healthcare Data Analytics- Data and Information – Types of Data Analytics -Components of Healthcare Analytics – Business Intelligence Applications in Healthcare – Electronic Health Records- Components of EHR – Administrative System Components -Clinical Documentation - Coding Systems- Benefits of EHR- Barrier to Adopting EHR- Challenges – Phenotyping Algorithms. Case Study: To analyze the impact of EHR in Health Data Analysis.

Unit -2 Biomedical Image Analysis   L-9 Hours

Introduction-Biomedical Imaging Modalities- Computed Tomography-Positron Emission Tomography-Magnetic Resonance Imaging-Ultrasound-Microscopy-Biomedical Imaging Standards and Systems-Image Segmentation-Thresholding-Watershed Transform-Clustering-Image Registration-Registration Transforms-Similarity and Distance Metrics-Registration Optimizers-Feature Extraction-Object Features-Feature Selection and Dimensionality Reduction-Principal Component Analysis. Case Study: CT Image Analysis through Morphological operations.

Unit-3 Biomedical Signal Analysis   L-9 Hours

Introduction-Types of Biomedical Signals-Electroneurogram (ENG)- Electromyogram (EMG)- Electrocardiogram (ECG)- Denoising for a Single-Channel ECG-Denoising for a Multichannel ECG Power Line Interference- Electroencephalogram (EEG)- Electrogastrogram (EGG)- Phonocardiogram (PCG)- Other Biomedical Signals-The Least-Mean Squares (LMS) Algorithm-The Adaptive Recurrent Filter (ARF)- Denoising of Signals. Case Study: To analyze effect of filters in Denoising signals.

Unit-4 Data Mining for Clinical Text   L-9 Hours

Natural Language Processing- Report Analyzer- Text Analyzer- Core NLP Components-Morphological Analysis- Lexical Analysis- Syntactic Analysis- Semantic Analysis- Data Encoding- Mining Information from Clinical Text-- Information Extraction- Preprocessing-Context-Based Extraction- Extracting Codes- Current Methodologies- Rule-Based Approaches- Clinical Text Corpora and Evaluation Metrics- Clinical Applications- General Applications- EHR and Decision Support- Surveillance. Case Study: Perform word analysis on clinical text.

Unit-5 Advanced Data Analytics for Healthcare   L-9 Hours

Introduction to Visual Analytics and Medical Data Visualization- Clinical Data Types - Standard Techniques to Visualize Medical Data- Standard Techniques to Visualize Medical Data- High-Dimensional Data Visualization- Visualization of Imaging Data-Visual Analytics in Public Health and Population Research- Geospatial Analysis-Temporal Analysis- Beyond Spatio-Temporal Visualization- Visual Analytics for Clinical Workflow- Visual Analytics for Patients- Case Study: To analyze and visualize Spatial and Geospatial data.

Total : 45 Hours

G. Laboratory Experiments   **Total: 30 Hours**

Part A

TASK 1   To perform data import/export (.CSV, .XLS, .TXT) and data preprocessing operations on Clinical Data.

Tools: Python, Open CV

TASK 2   Implement the program to extract different attributes of the medical image.
Tools: Python, Pandas, Numpy, Matpolib, Open CV

TASK 3   Implement the program to improve the quality of the image by performing sharpening and smoothing operation.
Tools: Python, Numpy, Open CV

TASK 4   Implement region based image segmentation method to extract necessary information and remove the background details based upon different threshold values.
Tools: Python, Scikit, Matplotlib, Open CV

TASK 5   Identify an appropriate noise removal technique to remove the noise present in the given biomedical signal by comparing it with various noise removal filters.
Tools: Python, Numpy, Scipy, Open CV.

TASK 6   Implement the program for Pattern Classification using ECG signals.
Tools: Python, Numpy, Scipy, Open CV

TASK 7   To create and train your own text mining technique for performing key phrase extraction.
Tools: Python, Yake, spaCy, Open CV

TASK 8   Implement the program for Morphological analysis of Clinical text.
Tools: Python, Numpy, Open CV

TASK 9   Perform Temporal Analysis and apply the various visualization techniques.
Tools: Python, Pandas, Numpy, Matplotlib

TASK 10 Perform exploratory data analysis (EDA) on heart disease data set. Export all your vital signs as a dataset, import them inside a pandas data frame, visualize them and get different insights from the data.
Tools: Python, Tableau, Pandas, Numpy, Matplotlib

Part B (Use Cases)

Use Case 1: Analyze the impact of image preprocessing techniques on real time data.

Use Case 2: Analyze the effect of segmentation algorithms.

Use Case 3: Analyze effect of filters in Denoising signals.

Use Case 4: Extract important features from heart disease dataset.

Use Case 5: Analyze and visualize the medical data.

Laboratory: 30 Hours

H. Learning Resources (in IEEE Format)

i. Text Books:

1. Reddy, C. K., & Aggarwal, C. C.," Healthcare data analytics" (Vol. 36). CRC Press. 2015.

ii. Reference Books:

1. Nilanjan Dey et al, "Healthcare Data Analytics and Management", First Edition, Elsevier, 2018.
2. Chandan K. Reddy et al, "Healthcare Data Analytics", First Edition, CRC Press, 2020.

iii. Online References:

1. Health Care Analytics. Available: https://www.ibm.com/in-en/watson-health/learn/healthcare-analytics.
2. "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining". Accessed on April 2022 [online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6023432/ .
3. "Methodologies for designing healthcare analytics solutions: A literature analysis". Accessed on April 2022 [online]. Available: https://journals.sagepub.com/doi/10.1177/1460458219895386/

H. Learning Resources (in IEEE Format)

i. Text Books:

2. Reddy, C. K., & Aggarwal, C. C.," Healthcare data analytics" (Vol. 36). CRC Press. 2015.

ii. Reference Books:

3. Nilanjan Dey et al, "Healthcare Data Analytics and Management", First Edition, Elsevier, 2018.
4. Chandan K. Reddy et al, "Healthcare Data Analytics", First Edition, CRC Press, 2020.

iii. Online References:

4. Health Care Analytics. Available: https://www.ibm.com/in-en/watson-health/learn/healthcare-analytics.
5. "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining". Accessed on April 2022 [online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6023432/ .
6. "Methodologies for designing healthcare analytics solutions: A literature analysis". Accessed on April 2022 [online]. Available: https://journals.sagepub.com/doi/10.1177/1460458219895386/

COURSE CODE	COURSE TITLE	L	T	P	C
20232CS214	Time series and Forecasting	3	0	2	4

A. Preamble

This course will provide a detailed knowledge in time series analysis. The topics include exploratory data analysis, storage of temporal data, ARMA/ARIMA models and some advanced models, Machine learning and Deep learning techniques to classify and cluster time-series data

B. Prerequisite Courses
Nil

C. Course Objectives
Learners are exposed to
- Equip students with various forecasting techniques and knowledge on modern statistical methods for analyzing time series data.
- Understand violation of classical linear model assumptions when continuous random variables changeover time.
- Learn about theoretical and empirical solutions when assumptions are violated.

D. Course Outcomes
Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Identify the elimination techniques to remove trend and seasonality in time-series analysis.	K2
CO2	Examine exploratory analysis on time series dataset for data visualization.	K3
CO3	Use the various time series models to solve problems.	K3
CO4	Implement suitable Machine Learning algorithms on time-series data.	K3
CO5	Apply suitable metrics to measure performance for the time series models.	K3
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of COs with Program outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	2	2					2	2
CO2	3	2	2		2		2	2
CO3	2	3			2		2	2
CO4	2	2	3		3		3	3
CO5	2	2	2		3		3	3
H- Strong; M-Medium; L-Low								

F. Course Contents

Unit-1 Introduction to Time Series Analysis   L-9 Hours

Examples and Objectives of Time Series Analysis – Time series models: zero-mean model – trends – seasonality – harmonic regression – Smoothing with a finite moving average filter – Exponential smoothing – Smoothing by elimination of high-frequency components – Trend elimination by differencing

Unit -2 Exploratory Data Analysis, Simulation, and Storing   L-9 Hours

Augmented Dickey–Fuller test - Rolling windows – Autocorrelation - Partial Autocorrelation - Spurious Correlations – Simulating Time Series Data: Simulation Versus Forecasting - Monte Carlo Simulations - Statistical Simulations - Deep Learning Simulations - Storing Temporal Data: SQL - NoSQL - InfluxDB – Prometheus.

Unit-3 Stationary and Non-Stationary Time Series Model   L-9 Hours

ARMA model - The Durbin–Levinson Algorithm - The Innovations Algorithm - The Periodogram – ARIMA – SARIMA - Multivariate AR Processes - State-Space Models - The ARAR algorithm - The Holt–Winters Algorithm - The Holt–Winters Seasonal Algorithm.

Unit-4 Machine Learning for Time Series Data   L-9 Hours

Generating and Selecting Features for a Time Series, General Considerations When Computing Features: The Nature of the Time Series – Streaming data- Domain Knowledge - External Considerations – Selecting and Generating Features for Classification - Gradient boosted trees - Generating Clustering Features from the Data - Temporally Aware Distance Metrics for Clustering: Fréchet distance - Pearson correlation - Longest common subsequence

Unit-5 Performance Analysis   L-9 Hours

Measuring Error: How to Test Forecasts - Model-Specific Considerations for Back testing - Estimating Uncertainty - Predicting Multiple Steps Ahead - Recursive Approach to Distant Temporal Horizons - Multitask Learning Applied to Time Series - Model Validation Gotchas - Performance Considerations in Fitting and Serving Time Series Models – Time series Applications.

Total : 45 Hours

G. Laboratory Experiments   Total: 30 Hours
Part A

TASK 1   The time series data set has some fluctuations/aberrations. Apply exponential smoothing technique to the data set.

Tools: python: numby, math

TASK 2   Consider the data set given to you. It has some trend and seasonal components. Apply filter to this time series data and discuss the results.
Tools: python : numby, matplotlib

TASK 3   The given time series data set has trend components. Remove the trend using differencing technique.
Tools: python: numpy, math

TASK4   Check if a particular time-series data is stationary or not. Apply Augmented Dickey–Fuller test using python language. Apply rolling windows for the given time-series data using python language.
Tools: python: pandas

TASK5   Plot the autocorrelation function for the given set of time-series data using python language. Use partial autocorrelation for the given data as well.
Tools: python: math, pandas

TASK6   Create a simulation of the time-series data using python language. Also, read and write time-series data to InfluxDB database.
Tools: python: pandas, TensorFlow

TASK7   Implement Autoregressive models and Moving Average models to the given time-series data
Tools: python: TensorFlow, keras

TASK8   Compare the results produced by Autoregressive Integrated Moving Average model and Vector Autoregression model.
Tools: python: numpy, math

TASK9   Write a program that uses the Durbin-Levinson algorithm to do time series analysis.
Tools: python: numpy

TASK10   Consider the time series data set given to you. Do stationary analysis using the innovations algorithm.
Tools: python: numpy, math

TASK11   For the given data set, choose which model is suitable to use: ARMA, ARIMA, SARIMA.
Tools: python: pmdarima

TASK12   Do multivariate time series analytics using ARAR algorithm for the given data set.
Tools: python: matplotlib, pandas, statsmodel

TASK13   Apply both the Holt-Winters algorithm and the Holt-Winters Seasonal algorithm to the datasets given to you.
Tools: python: matplotlib, pandas, statsmodel

TASK14   Check the performance of different distance metrics on clustering of time-series data
Tools: python: pandas, math

TASK15   Apply ANN to the given time-series data. Compare its performance.
Tools: python: TensorFlow, numpy

Part B (Use Cases)

Use case 1: Web traffic forecasting

Use case 2: Stock prices forecasting

Use case 3: Demand and sales forecasting

Use case 4: Individual household electric power consumption.

Use case 5: Anomaly detection for fraud detection

Laboratory: 30 Hours

H. Learning Resources (in IEEE Format)

i. Text Books:

1. Peter J. Brockwell, Richard A. Davis, "Introduction to Time Series and Forecasting", Second Edition, Springer, 2016 (UNIT I, III).

2. Aileen Nielsen, "Practical Time Series Analysis - Prediction with Statistics and Machine Learning", O'Reilly publications, First Edition, 2019. (UNIT II, IV, V).

ii. Reference Books:

1. Robert H. Shumway, David S. Stoffer, "Time Series Analysis and Its Applications - With R Examples", Fourth Edition, Springer, 2016.

2. William.W.S.Wei, "Time Series Analysis – Univariate and Multivariate Methods", Second Edition, Pearson, 2006.

iii. Online References:

1. Pandas python library - Accessed on 08 October 2022 [Online] Available: https://pandas.pydata.org/

2. Scikit-learn ML library- Accessed on 04 October, 2022 [Online]Available: https;//scikit-learn.org/stable.

COURSE CODE	COURSE TITLE	L	T	P	C
20232CS102	**Data Privacy and Security**	**2**	**0**	**0**	**2**

A. Preamble

This course is to familiarize the students with Basic concepts in security and privacy. This course addresses the security and privacy issues in legacy systems and also covers privacy policies laws, regulations, access control, monitoring, auditing, security information and event management..

B. Prerequisite Courses

20232CS206 - Cryptography and Network Security

C. Course Objectives

Learners are exposed to
- Familiarizes with the basic concepts of privacy in today's environment.
- Demonstrate a detailed of Privacy and Storage security and related Issues.
- Explain the basic concepts of Identity, cloud policy and Access Management.
- Develop the security essentials in IT Sector.
- Know the Monitoring and auditing the applications.

D. Course Outcomes

Upon the successful completion of the course, students will be able to:

CO No's	Course Outcomes	K - Level
CO1	Understand the concepts of Data privacy and Privacy Policies	**K2**
CO2	Explain Privacy enabling technologies and Storage Area Network security	**K2**
CO3	Apply the Access control mechanisms based on role and identity.	**K3**
CO4	Plan and Organize the protection of data and information system	**K3**
CO5	Illustrate the proactive monitoring and audit the information systems	**K2**
Knowledge Level (Based on revised Bloom's Taxonomy) K1-Remember K2-Understand K3-Apply K4-Analyze K5-Evaluate K6-Create		

E. Correlation of COs with Program outcomes and Programme Specific Outcomes:

Cos	PO1	PO2	PO3	PO4	PO5		PO10	PO11	PSO1	PSO2
CO1	2	2	3							
CO2	3	2	2	2						2
CO3	2	2	2	3					3	2
CO4	2	2	2							2
CO5	3	2	2							2
H- Strong; M-Medium; L-Low										

F.   Course Contents

Unit 1 Fundamental of Data Privacy   **L-6   Hours**

Introduction- Data Privacy Concepts, Data Privacy Attacks, Data linking and profiling, access control models, role-based access control, privacy policies, their specifications, languages and implementation, privacy policy languages, privacy in medical and financial domains.

Unit 2 Privacy and Storage   **L-6 Hours**

Privacy on the Internet - Privacy Enhancing Technologies - Detection of Conflicts in security policies - privacy and security in environment monitoring systems. Storage Area Network Security - Storage Area Network Security Devices - Risk management - Physical Security Essentials.

Unit 3 Access Control and Identity Management   **L-6   Hours**

Access control requirements for Cloud infrastructure - User Identification - Authentication and Authorization - Roles-based Access Control - Multi-factor authentication - Single Sign-on, Identity Federation - Identity providers and service consumers - Storage and network access control options - OS Hardening and minimization - Verified and measured boot.

Unit 4 Security Management   **L-6 Hours**

Data Protection strategies: Data retention, deletion and archiving procedures for tenant data, Encryption, Data Redaction, Tokenization, Obfuscation, PKI and Key. Components of an Information System, Securing the Components, Balancing Security and Access, Information security essentials for IT Managers.

Unit 5 Monitoring, Auditing and Management   **L-6 Hours**

Proactive activity monitoring - Incident Response, Monitoring for unauthorized access, malicious traffic, abuse of system privileges - Events and alerts - Auditing – Record generation, Reporting and Management, Tamper-proofing audit logs, Quality of Services, Secure Management, User management, Identity management, Security Information and Event Management.

Total : 30 Hours

G.   Learning Resources (in IEEE Format)

i.   Text Books:

1. Brij B. Gupta , Dharma P. Agrawal, H. Wang "Computer and Cyber Security: Principles, Algorithm, Application, and Perspectives", CRC Press, 2022 (Unit 1,2)
2. Michael E. Whitman, Herbert J. Mattord, Principles of Information Security, Seventh Edition, Cengage Learning, 2021 (Unit 3, 4 & 5)

ii.   Reference Books:

1. Siani Pearson, George Yee "Privacy and Security for Cloud Computing", Computer Communications and Networks, Springer, 2021.
2. Mather, Kumaraswamy and Latif, Cloud Security and Privacy!, OREILLY 2021

iii.   Online References:

1. "Misinformation and Privacy" Accessed on Apr. 20, 2022 [Online]. Available: https://nptel.ac.in/courses/106106146.
2. "Data Privacy and Fundamental". Accessed on: Apr. 15, 2021 [Online]. Available: https://www.classcentral.com/course/northeastern-data-privacy-20025.