Vel Tech
Rangarajan Dr. Sagunthala
R&D Institute of Science and Technology
(Deemed to be University Estd. u/s 3 of UGC Act, 1956)
Chennai, Tamil Nadu
India

35th BOS Meeting

with effect from
26.03.2022

M.Tech - Big Data Analytics

M.Tech (R16) - 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.

M. Tech BIG DATA ANALYTICS
CBCS VTUR16

Honors / Specialization / Minor

(With effect from 2021-2022)

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
Foundation Courses (FC)	4
Program Core (PC)	30
Program Elective (PE)	12
Independent Learning(IL)	8
Project Work	26
TOTAL	80

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

Department of Computer Science and Engineering

M. Tech Big Data Analytics - CBCS VTUR16

S. No	Subject Code	CBCS R-16		L	T	P	C	PG.NO
Foundation Course								
1	2160MA102	Probability and Statistics	FC	4	0	0	4	8
Program Core								
2	2161CS209	Advanced Data Structures and Algorithms	PC	3	0	2	4	12
3	2161CS210	Modern Operating Systems	PC	3	0	2	4	15
4	2161CS217	Advanced Data Base Management Systems	PC	3	0	2	4	18
5	2161CS130	Parallel Computer Architecture	PC	3	0	0	3	22
6	2161CS211	Software Engineering Principles and Practices	PC	3	0	2	4	24
7	2161CS218	Advanced Computer Network	PC	3	0	2	4	27
8	2161CS129	Machine Learning Techniques	PC	3	0	0	3	32
9	2161CS304	Machine Learning Techniques Laboratory	PC	0	0	2	1	35
10	2161CS140	Artificial Intelligence	PC	3	0	0	3	37
			Total Credits				**30**	
Program Specific Elective								
Data Science								
1	2162CS201	Foundation of Data Science	PE	2	0	2	3	41
2	2162CS202	Big Data Analytics	PE	2	0	2	3	44
3	2162CS203	Data Visualization	PE	2	0	2	3	47
4	2162CS205	Deep Learning	PE	2	0	2	3	50
5	2162CS210	Healthcare Analytics	PE	2	0	2	3	53
6	2162CS211	Time series and Forecasting	PE	2	0	2	3	57

Independent Learning				
1	2163MG401	Research Methodology	Independent Learning	2
2	2163CS402	ICT Tools	Independent Learning	2
3	2163CS501	Research Seminar	Independent Learning (Anyone)	2
4	2163CS801	Field Study		
5	2163CS802	Internship		
6	2163GE401	Business Communication and Technical Writing	Independent Learning	2
Project Work				
1	2164CS601	Project Phase 1	Project Work	10
2	2164CS701	Project Phase 2	Project Work	16

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

FOUNDATION COURSE

COURSE CODE	COURSE TITLE	L	T	P	C
2160MA102	PROBABILITY AND STATISTICS	3	2	0	4

A. Preamble

To provide an in-depth knowledge about Probability, Sampling, Correlation and Random process

B.Prerequisite Courses

Concrete Mathematics

C.Course Objectives

- Identify the basic concepts of probability and functions.
- Apply graphical methods for Sampling distribution with variance and standard deviations.
- Apply concepts of various Hypothesis testing along with Chi-square test
- Calculate Correlation and Regression and Apply Random Process Classifications

.

D.Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Understanding basic concepts of probability	K2
CO2	Applying basic concepts of different types of sampling	K3
CO3	Applying basic concepts of Hypothesis testing and Chi-square testing to solve simple problems	K3
CO4	Applying basic concepts of Correlation	K3
CO5	Applying basic 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	M			M			M	M
CO2	M			M			M	M
CO3	M			M			M	M
CO4	M			M			M	M
CO5	M			M			M	M

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Probability Function L-9 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.

UNIT2 Distribution Function L-9 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-9 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-9 Hours

Correlation – Scatter diagram – Karl Pearson 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-9 Hours

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

Total: 45 Hours

G.Learning Resources (in IEEE Format)

i.Text Books:

1. Richard A. Johnson and C. B. Gupta, Probability and Statistics for Engineers, (7th Ed.), Pearson Education, Indian Impression – 2006.

ii.Reference Books:

1. Probability & Statistics by T.K.V.Iyengar & B.Krishna Gandhi & Others, S.Chand, 2012
2. Probability & Statistics by William Mendenhall & Others, Cengage Publications.

iii.Online References:

1. https://www.cimt.org.uk/projects/mepres/alevel/fstats_ch4.pdf
2. https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/
3. https://oli.cmu.edu/courses/probability-statistics-open-free/

PROGRAM CORE

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS209	ADVANCED DATA STRUCTURES AND ALGORITHMS	3	0	2	4

A. Preamble

To provide an in-depth knowledge in problem solving techniques and data structures

B . Prerequisite Courses

Data Structure

C. Course Objectives

Students undergoing this course are expected to

- To develop a logic for graphical modelling of the real life problems.
- To suggest appropriate data structure and algorithm for graphical solutions of the problems.
- To understand advanced data structures to solve complex problems in various domains.
- To build the logic to use appropriate data structure in logical and computational solutions.
- To understand various algorithmic strategies to approach the problem solution

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 concepts of trees structures, Hashing	K2
CO2	Demonstrate the depth knowledge of Disjoint sets	K3
CO3	Illustrate some of the advanced algorithms in graphs	K2
CO4	Demonstrate the depth knowledge of NP completeness and string matching algorithms.	K3
CO5	Demonstrate knowledge of Randomized Algorithms	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	M						M	
CO2			H	M			M	
CO3	M		M	H			M	
CO4	M		H	M			M	
CO5	M	M	H	M			M	

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Elementary Data Structures L-9 Hours

Probabilistic Analysis, Amortized Analysis, Competitive Analysis, Hash tables: Direct Addressing, Functions and tables, Open addressing, perfect Hashing , ed-Black Trees, Splay Trees, Augmented Data Structures

UNIT 2 Advanced Data Structures L-9 Hours

B-trees, Fibonacci heaps, Van Emde Boas Trees, Data Structures for Disjoint Sets.

UNIT 3 Graphs & Algorithms L-9 Hours

Representation, Maximum Flow-Flow networks, The Ford-Fulkerson method, maximum bipartite matching, Push-relabel algorithms, the relabel-to-front algorithm.

UNIT 4 String Matching And Approximation Algorithms L-9 Hours

String Matching Algorithms: Rabin-Karp, Knuth-Morris-Pratt, String Matching with finite automata-Approximation algorithms: Need of approximation, Introduction to P, NP, NP-Hard and NP-Complete; Vertex Cover problem, TSP, Set Cover, Subset sum Problem.

UNIT 5 Randomized Algorithms L-9 Hours

Introduction, Type of Randomized Algorithms, Min- Cut, 2- SAT; Game Theoretic Techniques, Random Walks.

Total: 45 Hours

B. Laboratory Experiments Total: 30 Hours

LIST OF EXPERIMENTS

Tasks	Experiments
Task 1	Create a hash table and perform open addressing for a set of values.
Task 2	For a given graph perform graph coloring.
Task 3	Create a Btree for a given set of values and perform insert and delete operations.
Task 4	Perform two types of rotations in Red Black Tree.
Task 5	Find out the Maximum flow and Minimum cut in a graph for any given number of nodes.
Task 6	Implement string matching by Rabin-Karp Algorithm.

Task 7	Implement string matching by Knuth-Morris-Pra Algorithm.
Task 8	Solve TSP for a given graph.
Task 9	Solve Set Cover Problem for given set of elements.
Task 10	Write a program to solve subset sum.

Software: C / C++ / JAVA

Total: 75 Hours

C. Learning Resources (in IEEE Format)

i. Text Books:

1. Thomas Coremen, "Introduction to Algorithms", Third edition, Prentice Hall of India (2009).
2. Motwani R., Raghavan P., "Randomized Algorithms", Cambridge University Press, 1995.

ii. Reference Books:

1.Kleinberg J., Tardos E., "Algorithm Design", 1st Edition, Pearson, 2012.
2.Vazirani, Vijay V., "Approximation Algorithms", Springer, 2001.
3.Floyds L.R,"Graph Theory Applications ", Springer, 1994.

iii. Online References:

1. https://www.javatpoint.com/daa-rabin-karp-algorithm
2. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-851-advanced-data-structures-spring-2012/index.htm
3. https://www.cs.cmu.edu/~ckingsf/class/15351-f14/resources.html

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS210	MODERN OPERATING SYSTEMS	3	0	2	4

A. Preamble

To provide an in-depth knowledge in operating system concepts.

B. Prerequisite Courses

Microprocessor and Microcontrollers

C. Course Objectives

- To gain knowledge on Distributed operating system concepts that includes architecture, Mutual exclusion algorithms, Deadlock detection algorithms and agreement protocols
- To gain insight on to the distributed resource management components viz. the algorithms for implementation of distributed shared memory, recovery and commit protocols
- To know the components and management aspects of Real time, Mobile operating Systems

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Understanding the various types of operating systems and synchronization problems.	K2
CO2	Differentiate Distributed Operating system with Traditional OS	K3
CO3	Demonstrate the concepts of Deadlocks in distributed OS	K3
CO4	Conceptualize of failure and fault tolerance issues in Distributed OS.	K3
CO5	Analyze the design issues of Real time OS.	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	PO6	PSO1	PSO2
CO1	M		M				M	M
CO2	M		M				M	M
CO3	M		M	M			M	M
CO4	M		M	M	M		M	M
CO5	M	H	M	M	M		M	M

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Fundamentals of Operating Systems L-9 Hours

Overview – Synchronization Mechanisms – Processes and Threads - Process Scheduling– Deadlocks: Detection, Prevention and Recovery – Models of Resources – Memory Management Techniques.

UNIT 2 Distributed Operating System L-9 Hours

Scheduling: Issues in load distributing, Components of load distributing algorithms, Stability, Load distributing algorithms, Performance Comparison, Selecting load sharing Algorithm-Synchronization: Physical and logical clocks -Distributed Mutual Exclusion: Mutual Exclusion algorithms and its classification

UNIT 3 Distributed Deadlock Handling L-9 Hours

Introduction, deadlock handling strategies, detection: Issues and resolution, Control Organizations, Centralized algorithms, Distributed algorithms, Hierarchical algorithms.

UNIT 4 Distributed Fault Handling L-9 Hours

Agreement Protocol: System Model, Classification, solution to Byzantine Agreement Problem- Fault Recovery: Concepts, Classification of failures, Backward error recovery, Recovery in concurrent Systems, Consistent Check Points, Synchronous and Asynchronous check pointing and recovery-Fault tolerance: Issues, Atomic actions and committing, Commit Protocols, Non-blocking Commit protocols, Voting protocols and Dynamic Voting Protocol.

UNIT 5 Real Time Operating Systems L-9 Hours

Types of Real time tasks, Timing Constraints, Modeling Timing Constraints-Task Scheduling: Types of tasks and their characteristics, Clock driven Scheduling, Hybrid Schedulers, Event driven Scheduling, EDF Scheduling, Rate Monotonic Algorithm - Resource Handling: Resource Sharing, Priority Inversion, PIP, PCP, HLP.

Total: 45 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Mukesh Singhal, Niranjan G. Shivaratri, "Advanced Concepts in Operating Systems: Distributed, Database and Multiprocessor Operating Systems", McGraw Hill education, 2017.
2. C.M. Krishna and K.G. Shin, “Real-Time Systems: Theory and Practice”, McGraw Hill Education; 1st edition (1 July 2017)

ii. Reference Books:

1. Abraham Silber Schatz, Peter Baer Galvin, Greg Gagne, “Operating System Concepts”, 10th Edition, John Wiley & Sons, 2018.
2. Andrew Tanenbaum, “Modern Operating System”, Pearson Edition, 4th Edition, 2016.

iii. Online References:

1. https://quescol.com/distributed-system/
2. https://www.udacity.com/course/introduction-to-operating-systems--ud923
3. https://www.udacity.com/course/advanced-operating-systems--ud189

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS217	Advanced Data Base Management Systems	3	0	2	4

A. Preamble

To provide an in-depth knowledge about how database systems are maintained, manage and mitigate the current real-world scenarios via distributed databases, multi version concurrency control, summarizing and provide statistic-based decision-making with huge databases like warehouses, big data. This course also provides a basic mining terminologies like rule mining and clustering.

B. Prerequisite Courses

Data Base Management System

C. Course Objectives

Students undergoing this course are expected to

- Describe the fundamental elements of relational database management systems
- Explain the basic concepts of relational data model, entity-relationship model, relational database design, relational algebra and SQL.
- Design ER-models to represent simple database application scenarios
- Convert the ER-model to relational tables, populate relational database and formulate SQL queries on data.

D. Course Outcomes

CO Nos.	Course Outcomes	K - Level
CO1	Infer the basic OO concepts underlying databases	K2
CO2	Relate object-oriented databases and ODMG model	K2
CO3	Explain the concurrency control through parallel and distributed databases	K2
CO4	Describe the decision-making process in warehouse and through mining process	K2
CO5	Interpret modern data models relating to real time scenarios	K2

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

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

H- Strong; M-Medium; L-Low

F. Course Content:

UNIT 1 Review of Relational Data Model and Oodbs L -9 Hours

Relational model concepts; Relational model constraints and relational database schemas; Update operations, anomalies, dealing with constraint violations, Types and violations. 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- 9 Hours

Overview of OOP; Complex objects; Identity, structure etc. Object model of ODMG, Object definition Language ODL; 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, The nested relational model. Overview of C++ language binding; Mongoose- Object relational features-CRUD operations in Mongoose environment;

UNIT 3 Parallel and Distributed Databases L-9 Hours

Architectures for parallel databases; Parallel query evaluation; Parallelizing individual operations; Parallel query optimizations; 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, Shading; Distributed transactions; Distributed Concurrency control and Recovery; MVCC with PostgreSQL and SQLite.

UNIT 4 Data Warehousing, Decision Support and Data Mining L-9 Hours

Introduction to decision support; OLAP, multidimensional model; Window queries in SQL; Finding answers quickly; 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; ROC and CMC Curves; Clustering; Similarity search over sequences; Incremental mining and data streams; Additional data mining tasks.

UNIT 5 Enhanced Data Models for Some Advanced Applications L-9 Hours

Active database concepts and triggers; Temporal, Spatial, and Deductive Databases – Basic concepts. More Recent Applications: Mobile databases; Multimedia databases; Geographical Information Systems; Genome data management. NoSQL with Object Oriented Concepts- MongoDB-Mongoose; Graph databases- Neo4j- CRUD operation using Neo4j; Neo4j vs Graph; Query modelling using Graph; Handling Big Data using MongoDB-Map Reduce using MongoDB-Hadoop Node configuration;

Total: 45 hours

G. Laboratory Experiments

Total: 30 Hours

Laboratory Tasks

Part-A

Task-01: Designing databases using Postgres and SQLite

Task-02: Performing Equivalent queries using Oracle for simple clauses and operators like having, group by; like, is, as, in, not in; < ,> ; and, or;

Task-03: Performing equivalent queries for nested and join queries using join..with , join..in, using SQLite

Task-04: Reporting logs of listener controls using listctl of Oracle SQL

Task-05: Performing Multi version concurrency control using Snapshot using Oracle SQL

Task-06: Making Java GUI and performing OO transactions using Postgres and SQLite

Task-07: designing QBE design performing CRUD operations in postgres and SQLite queries.

Task-08: Devising Cursors for coherence control of database objects.

Task-09: Devising Cluster partitions using Oracle SQL for Schemas.

Task-10: Performing Cluster partitions or map reduce using MongoDB.

Task-11: Performing CRUD operations using MongoDB

Task-12: Performing CRUD operations using GraphQL/Neo4j

Part-B

Use Case-1: Building a Cart analysis for Myph

Myph has just launched their brand new phone range to the eager reception of the consumer market cart analysis. The product's data model has a unique menu that identifies the product, title, description, a stock quantity, and pricing information about the item. All products have categories. To be able to provide a list of all the products in a category, amend the data model with a collection of documents for each category and contain the path for that category in the category tree. Use cart analysis in developing different consumer selection options. Would this answer outlier selection in cart i.e., surplus selections? Is Relational database application can answer these transactions? How recovery is made through carting and commerce?

Use Case-2: Indexing various devices in IoT platform

A generic IoT platform required support for data from a wide range of devices, some of which could not be envisaged while developing the platform. The proficient work necessitated a data storage mechanism that could handle data from different types of devices. Indexing support makes it easy to pull data using a single index or multiple indexes such as device id with location id. Records for a

particular device in different locations are easily accessed. Common parameters like temperature from different types of devices and their records are retrieved fast through these indexes. How the application could lead to the choice of JSON-based document database, MongoDB? Assume or create JSON script in support of this.

Total: 75 hours

H. Learning Resources (in IEEE Format)

i. Text Book

1. J.D. Ullmann et al, "Database Systems: The Complete Book", Second Edition, Pearson Ed, Inc, 2009.

ii. Reference Book

1. 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]
2. Agus Kurniawin, “Python and SQLite Development”, First Edition, PE Press, January 2021. [unit-5]
3. Stephano Ceri and Giusepe Pelagatti,“Distributed databases: Principles & Systems”, Mc Graw Hill (India) Pvt Ltd, 2017

iii. Online Resource

1. “Designing local library models”, Accessed on: May 05, 2021[online]. Available: https://developer.mozilla.org/en-US/docs/Learn/Server-side/Express_Nodejs/mongoose.
2. Vivian Siahaan and Rismon Hasiholan Sianipur,” The fast way to learn Java GUIwith PostgreSQL and SQLite”, First Ed- E-book, Accessed on : October 27,2021[online].Available:https://www.google.co.in/books/edition/The_Fast_Way_to_Learn_Java_GUI_with_Post/uPvJDwAAQBAJ?hl=en&gbpv=1&dq=sqlite&printsec=frontcover.
3. Michael Hunger, Ryan Boyd & William Lyon, “The Definitive Guide to Graph Databases for the RDBMS Developer”, first Ed, E-book, Accessed on: October 27, 2021 [online & Download]. Available: https://neo4j.com/whitepapers/rdbms-developers-graph-databases-ebook/

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS130	PARALLEL COMPUTER ARCHITECTURE	3	0	0	3

A. Preamble

To provide an in-depth knowledge about parallelism, scalability in computer architecture.

B. Prerequisite Courses

Computer Organization and Architecture

C. Course Objectives

- Able to Describe current approaches to Parallel Computing
- Explain the design principles of the hardware support for the shared memory and message passing programming models
- To Implement Synchronization methods for shared memory and message passing parallel computers
- To Design scalable parallel software and analyze its performance

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 terminology and concepts of parallel Architecture	K1
CO2	Interpret the performance of Parallel Programming	K2
CO3	Analyze the Workload driven emulation for memory	K3
CO4	Familiarize with Snoop Based Multiprocessor Design	K2
CO5	Recognize the concept of Scalable Multiprocessors	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	M		M				M	M
CO2	M		M				M	M
CO3	M		M				M	M
CO4	M		M				M	M
CO5	M		M				M	M

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction L-9 Hours

Introduction, Parallel Architecture, Convergence of Parallel Architecture, Fundamental Design Issues, Parallel Application, Parallelization Process, Example Program.

UNIT 2 Performance Analysis L-9 Hours

Partitioning for Performance, Data Access and Communication in a Multi-Memory System, Orchestration for Performance, Case Studies.

UNIT 3 Workload Driven Evaluation L-9 Hours

Scaling Workloads and Machines, evaluating a Real Machine, Cache Coherence, Memory Consistency, Snooping Protocols, Synchronization

UNIT 4 Snoop Based Multiprocessor Design L-9 Hours

Correctness Requirements, Base Design, Multi-Level Cache Hierarchies, Split Transaction Bus, Case Studies

UNIT 5 Scalable Multiprocessors L-9 Hours

Scalability, Realizing Programming Models, Physical DMA, User Level Access, Dedicated Message Processing, Shared Physical Address Space, Clusters and Networks of Workstations

Total: 45 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. D.E. Culler, J.P. Singh, and A. Gupta, “Parallel Computer Architecture - A Hardware/Software Approach”, Morgan Kaufmann Publishers, 2010.

ii. Reference Books:

1. N.E. Jerger and Li-Shiuan Peh,”On-Chip Networks”, Morgan and Claypool, 2009.
2. D.J. Sorin, M.D. Hill, and D.A. Wood,”A Premier on Memory Consistency and Cache Coherence”, Morgan and Claypool, 2011.
3. John P. Hayes, “Computer Architecture and Organisation” ,MCGraw Hill. 3rd Edition, 1998,
4. Hwang K. and Briggs. F.A, “Computer Architecture and Parallel Processing”, MCGraw Hill, 1985.

iii. Online References:

1.http://15418.courses.cs.cmu.edu/spring2017/lectures

2.https://nptel.ac.in/courses/106/104/106104024/

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS211	SOFWARE ENGINEERING PRINCIPLES AND PRACTICES	3	0	2	4

A. Preamble

To provide an in-depth knowledge about the Software Architecture, Software Documentation.

B. Prerequisite Courses

Database Management Systems

C. Course Objectives

- To give a broad overview of the software lifecycle and the development process.
- To provide students with an understanding of how software engineering helps to manage software development.

D. Course Outcomes

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

CONo's	Course Outcomes	K - Level
CO1	Understand the basics of software engineering.	K2
CO2	Demonstrate use of software configuration and quality management.	K2
CO3	Execute various models used in software lifecycle.	K3
CO4	Understand various design and testing techniques used in software	K2
CO5	Apply various tools and techniques for software maintenance.	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	M						M	
CO2		M						
CO3	H	M	M		H		M	M
CO4	M	M			H			M
CO5			H		H			

H- High; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction to Software Engineering L-9 Hours

What Is Software Engineering, Phases in the Development of Software, Software Engineering Ethics, planning a Software Development Project, Controlling a Software Development Project, The Waterfall Model, Agile Methods, The Rational Unified Process, Model-Driven Architecture.

UNIT2 Configuration and Software Quality Management L-9 Hours

Tasks and Responsibilities, Configuration Management Plan, People Management, Team Organization, On Measures and Numbers, Taxonomy of Quality Attributes, Perspectives on Quality, The Quality System, Software Quality Assurance, Capability Maturity Model, Algorithmic Models, Guidelines for Estimating Cost, Distribution of Manpower over Time, Agile Cost Estimation, A Taxonomy of Software Development Projects, Risk Management.

UNIT 3 Software Life Cycle L-9Hours

Requirements Elicitation, Requirements Documentation and Management, Requirements Specification Techniques, Verification and Validation, Classic Modeling Techniques, Unified Modeling Language, Software Architecture and the Software Life Cycle, Architecture Design, Architectural Views, Architectural Styles, Software Architecture Assessment.

UNIT 4 Software Design and Testing L-9 Hours

Design Considerations, Classical Design Methods, Object-Oriented Analysis and Design Methods, How to Select a Design Method, Design Patterns, Design Documentation, Verification and Validation, Test Objectives, Testing and the Software Life Cycle, Verification and Validation Planning and Documentation, Manual Test Techniques, Coverage-Based Test Techniques, Fault-Based Test Techniques, Error-Based Test Techniques.

UNIT 5 Software Maintenance and Tools L-9 Hours

Maintenance Categories Revisited, Major Causes of Maintenance Problems, Reverse Engineering and Refactoring, Toolkits, Language-Centered Environments, Integrated Environments and Work Benches, Process-Centered Environments, User Interface Design, Human Factors in Human–Computer Interaction, Role of Models in Human–Computer Interaction, Design of Interactive Systems, Reuse Dimensions, Reuse and the Software Life Cycle, Reuse Tools and Techniques, Component-Based Development Process and Component Life Cycle, Service-Oriented Software Engineering, Challenges of Global System Development.

Total: 45 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Hans van Vliet, "Software Engineering: Principles and Practice", 3rd Edition, 2010.

ii. Reference Books:

1. Deepak Jain, "Software Engineering-Principles and Practices", Oxford University Press, First Edition, 2018.

iii. Online References:

1.https://lecturenotes.in/subject/121/principles-and-practices-in-software-engineering-ppse/note

2. https://www.sei.cmu.edu/education-outreach/courses/course.cfm?coursecode=P35

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS218	ADVANCED COMPUTER NETWORKS	3	0	2	4

A. Preamble

The course deals with the concepts of networking, Modeling and Simulation techniques. Also represents the working principles of wireless, mobile and satellite networks. It also covers the distributed system environment.

B. Prerequisite Courses

Computer Networks, Graph theory

C. Course Objectives

Students undergoing this course are expected to

- Be familiar with the concepts of TCP/IP reference model.
- Understand the concepts of Protocols, network interfaces.
- Performance study of Wireless networks and mobile networks.
- Identify the representation of Satellite networks.
- Understand the impact of Distributed systems.

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Develop the Working concepts of TCP-IP reference model and IPV6 message format and its services.	K3
CO2	Identify the networking concepts by Simulation and Modelling.	K3
CO3	Organize the Functionalities of wireless networks and mobile network.	K3
CO4	Build the scenario for Satellite Networks.	K3
CO5	Construct the Distribution System setup and its requirements.	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	M		M				M	M
CO2	M		M				M	M
CO3	M		M				M	M
CO4	M		M				M	M
CO5	M		M				M	M

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction L-9 Hours

Introduction, TCP/IP Fundamentals: TCP, UDP, IP, Performance Measurement of TCP/IP Networks: Reasons for Network Measure, Measurement Task, Classification of Measurement Tools Popular Measurement Tools and Their Applications.

UNIT 2 Network Simulation and Modeling L-9 Hours

The Role of Simulation, Steps of a Systematic Simulation Study, Types of Simulations, Simulation Validation and Verification, Confidence Level of Simulation Results, Simulation with Self-Similar Traffic, Classification of Simulation Tools, The ‘‘ns’’ Network, OPNET, TCP Modeling.

UNIT 3 Tcp/Ip Performance Over Wireless Networks and Mobile Networks L-9 hours

Wireless Networks, TCP Performance Issues Over Wireless Links, Improving TCP Performance over Wireless Links, Wireless System Evolution and TCP/IP, Cellular and Ad Hoc Networks, TCP Performance in Cellular Networks, TCP Performance in Ad Hoc Networks.

UNIT 4 Tcp/Ip Performance Over Satellite Networks L-9 Hours

A Brief History of Data Satellites, Motivations for Using Satellites, Types of Satellites Satellite Internet Architectures, Satellite Characteristics Affecting TCP, TCP Enhancements for Satellite Networks, Advanced Enhancements and New Versions of TCP, New Transport Protocols for Satellite Links.

UNIT 5 Distributed Systems L-9 Hours

Distributed System, Goals, Types of Distributed Systems, System Architectures, Architectures Versus Middleware, Self-Management in Distributed Systems,Processes:Threads, Virtualization, Clients, Servers, Code Migration, Communication:Fundamentals,Remote Procedure Call, Message-Oriented Communication, Stream-Oriented Communication, Multicast Communication.

Total: 45 Hours

G.Laboratory Experiments

Total: 30 Hours

TASK 1	Working with Networking Commands TCPdump, TCPstat, TTCP, Netperf, NetPIPE.
TASK 2	Extract the Packet from network interfaceand analysis the protocol, addressing. **Tools: Wireshark, JPCAP**
TASK 3	Perform simulation for Multi-Antenna Cellular Network. **Tools: NS3**
TASK4	Perform simulation for Heterogeneous Network. **Tools: NS3**
TASK5	Simulate the Under-Laid Cellular Networks in-terms of packet transfer and delay. **Tools: NS3**
TASK6	Simulate a geostationary satellite scenario with their functionalities. **Tools: NS2**
TASK7	Simulate a Low Earth Orbit (LEO) satelliteswith their workings. **Tools: NS3**
TASK8	Program to implement Chat Server. **Tools: C /Java**
TASK9	Program to implement Remote Procedure Call. **Tools: C/Java**
TASK10	To Simulate the Distributed Mutual exclusion. **Tools: C/Java**

Part-2

Use cases:

1.Geostationary satellite: Consider an INSAT-3A multipurpose geostationary satellite and two satellite terminals, one at Bangkok and the other at Baghdad. The position of Bangkok is 13.9-degree latitude north and 100.9-degree longitude east. The position of Baghdad is 33.8-degree latitude north and 44.4-degree longitude east. INSAT-3A is used to provide television broadcasting from Bangkok to Baghdad. INSAT-3A is positioned at 93.5 degrees longitude East. The traffic consists of a FTP source and a CBR stream. The simulation lasts for 50 secs.

Analyse the trace file and find the following:

a. End-to-end delay between two terminals
b. What is the nature of the delay that you expect in this scenario.

2.Low-earth orbit: Sets up two terminals, one in Boston and one at Berkeley. Send a packet in each second from Berkeley to Boston for whole day. Consider the following parameters for Iridium constellation:
The simulation lasts for one earth rotation (86400 sec).
Analyze the trace file and find the following:
i)What should be the nature of Hop-count for the packets sent from source to destination. Plot and verify this with graph.
ii)What should be the nature of end-to-end delay for the packets sent from source to destination and how it is related with the plot of count.

3.Chat server: Chat server is a standalone application that is made up the combination of two-application, server application (which runs on server side) and client application (which runs on client side). This application is using for chatting in LAN. To start chatting you must be connected with the server after that your message can broadcast to each and every client.

a. A simple server that will accept a single client connection and display everything the client says on the screen. If the client user types "bye", the client and the server will both quit.
b. A server as before, but this time it will remain 'open' for additional connection once a client has quit. The server can handle at most one connection at a time.
c. A server as before, but this time it can handle multiple clients simultaneously. The output from all connected clients will appear on the server's screen.
d. A server as before, but this time it sends all text received from any of the connected clients to all clients. This means that the server has to receive and send, and the client has to send as well as receive.
e. Wrapping the client from step 4 into a very simple GUI interface but not changing the functionality of either server or client. The client is implemented as an Applet, but a Frame would have worked just as well (for a stand-alone program).

4.Remote procedure call: Remote Procedure Call (RPC) is an inter-process communication that allows a computer program to cause a subroutine or procedure to execute in another address space (commonly on another computer on a shared network) without the programmer explicitly coding the details for this remote interaction.

a. The client calls the client stub. The call is a local procedure call, with parameters pushed onto the stack in the normal way
b. The client stub packs the parameters into a message and makes a system call to send the message. Packing the parameters is called marshalling.
c. The kernel sends the message from the client machine to the server machine.
d. The kernel on the server machine passes the incoming packets to the server stub.
e. Finally, the server stub calls the server procedure. The reply traces the same steps in the reverse direction

5. Distributed Mutual Exclusion:
Concurrent access of processes to a shared resource or data is executed in mutually exclusive manner. Only one process is allowed to execute the critical section (CS) at any given time

.In a distributed system, shared variables (semaphores) or a local kernel cannot be used to implement mutual exclusion. Message passing is the sole means for implementing distributed mutual exclusion.
Process1:Request resource:
Resource Allocated ◊ No more requests process for this resource.
Process2:RequestResource◊Denied
Process1:ExitResource:
Process2: Request Resource◊ Allocated
Press a key (except q) to enter a process into critical section. Press q at any time to exit. Process0enteredcriticalsection.
Error: Another process is currently executing critical section. Please wait till its execution is over. Process0exitscriticalsection.
Process1enteredcriticalsection.
Process1exitscriticalsection.
Process 2 entered critical section.
Error: Another process is currently executing critical section. Please wait till its execution is over.
Process 2 exits critical section.

Total: 75 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. High Performance TCP/IP: Networking Concepts, Issues, and Solutions, Mahbub Hassan and Raj Jain, IST Edition, 2009. (Unit 1-4).

2. Distributed Systems,Maarten Van Steen and Andrew S.Tanenbaum,Third Edition, Pearson Publication,2017.(Unit 5)

ii. Reference Books:

1.TCP/IP Illustrated (Volume I, Volume II and Volume III), W. Richard Stevens, Addison-Wesley.

iii. Online References:

1. TCP/IP Fundamentals, https://www.thegeekstuff.com/2011/11/tcp-ip-fundamentals/
2.NetworkModellingandSimulations, https://www.researchgate.net/publication/220674719_Network_modelling_and_simulation tools
3. TCP/IP Performance over wireless networkshttps://dl.acm.org/doi/10.1145/215530.554

4.TCP/IPPerformanceoverSatelliteCommunications, https://sites.cs.ucsb.edu/~ebelding/ courses /284/s06/papers/Satellite_TCP.pdf

5. Distributed systems, https://blog.stackpCath.com/distributed-system/

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS129	MACHINE LEARNING TECHNIQUES	3	0	0	3

A. Preamble

To provide an in-depth knowledge about machine learning concepts and identify applications suitable for different types of machine learning with suitable justification

B. Prerequisite Courses

Data Mining

C. Course Objectives

- Outline the basic concepts of machine learning
- Summarize supervised learning and classificationtechniques
- Apply the concept of unsupervised learning andClustering for applications
- Illustrate the concept of Dimensionality Reduction.
- Apply the theoretical and practical aspects of Machine learning Tools

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Develop an appreciation for what is involved in Learning models from data	K2
CO2	Understand a wide variety of learning algorithms	K3
CO3	Understand how to evaluate models generated from data	K3
CO4	Apply the algorithms to a real problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models	K3
CO5	Implement machine learning algorithms using various tools	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	M							
CO2	M				M		M	
CO3	M				M		M	
CO4	M				H		M	
CO5	M		M	H	H	M	H	

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction L-9 Hours

Machine Learning - Machine Learning Foundations - Types of Machine Learning – Design of a Learning system – Perspectives and Issues in Machine Learning – Concept Learning Task – Concept Learning as Search – Finding a Maximally Specific Hypothesis – Version Spaces and the Candidate Elimination Algorithm

UNIT 2 Supervised Learning L-9 Hours

Regression-Linear Regression- Classification - Naïve Bayes - Discriminant Functions -Probabilistic Generative Models -Probabilistic Discriminative Models-Decision Trees – Pruning –NeuralNetworks -Feed-forward Network Functions - Back propagation- Support vector machines - Ensemble methods-Bagging- Boosting.

UNIT 3 Unsupervised Learning L-9 Hours

Clustering- K-means - EM Algorithm -The Curse of Dimensionality - Dimensionality Reduction - Factor analysis – Linear Discriminant analysis-Principal Component Analysis - Probabilistic PCA

UNIT 4 Probabilistic Graphical Models L-9 Hours

Graphical Models - Undirected graphical models - Markov Random Fields - Directed Graphical Models -Bayesian Networks - Conditional independence properties - Inference – Learning Generalization - Hidden Markov Models

.

UNIT 5 Machine Learning Tools L-9 Hours

Machine Learning using Weka - Python for Machine Learning –Machine Learning using R-Introduction to Mahout.

Total: 45 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Tom M Mitchell, Machine Learning, First Edition, McGraw Hill Education, 2013.
2. Peter Flach, Machine Learning: The Art and Science of Algorithms that Make Sense of Data, First Edition, Cambridge University Press, 2012.

ii. Reference Books:

1. EthemAlpaydin, Introduction to Machine Learning 3e (Adaptive Computation and Machine Learning Series), Third Edition, MIT Press, 2014.
2. Jason Bell, Machine learning ,Hands on for Developers and Technical Professionals‖, First Edition, Wiley, 2014.
3. Stephen Marsland, Machine Learning – An Algorithmic Perspective, Second Edition, Chapman and Hall/CRC Machine Learning and Pattern Recognition Series, 2014.

iii. Online References:

1. https://warwick.ac.uk/fac/sci/dcs/teaching/modules/cs342/
2. https://nptel.ac.in/courses/106/105/106105152/
3. https://nptel.ac.in/courses/106/106/106106139/

COURSE CODE	COURSE TITLE	L	T	P	C
2161CS304	MACHINE LEARNING TECHNIQUES LABORATORY	0	0	2	1

A. Preamble

To understand the Image Processing technology and tool kits for programming.

B. Prerequisite Courses

C. Course Objectives

- Apply the tokenization using NLTK tool kit in python using any text file
- Demonstrate naïve Bayes in Mahout
- Apply the principles in decision tree and KNN for any dataset

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Interpret the programming concepts of lemmatization, tokenization	K3
CO2	Analyze the classification Process for real world example.	K3
CO3	Implement the clustering Process to identify the related entities.	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		M		H			M	
CO2	M	M		H	M		M	
CO3	M	M		H	M		M	

H- Strong; M-Medium; L-Low

F. Course Contents

Laboratory Experiments: **TOTAL:30 Hours**

Task 1 Perform a stop word removal a text file and print the text

Task 2 Implement lemmatization, tokenization using NLTK tool kit in python using any text file

Task 3 Perform a classification using naïve Bayes in Mahout

Task 4 Perform a classification using Support vector machine

Task 5 Implements k-means clustering algorithm

Task 6 Implement a decision tree and KNN for any dataset and compare the accuracy

Task 7 Illustrate PCA

Software Tools: R/ Python / Weka

G. Learning Resources (in IEEE Format)

i. Text Books:

1.Tom M Mitchell, Machine Learning, First Edition, McGraw Hill Education, 2013.
2. Peter Flach, Machine Learning: The Art and Science of Algorithms that Make Sense of Data, First Edition, Cambridge University Press, 2012.

ii. Reference Books:

1. Ethem Alpaydin, Introduction to Machine Learning 3e (Adaptive Computation and Machine Learning Series), Third Edition, MIT Press, 2014.
2.Jason Bell, Machine learning ,Hands on for Developers and Technical Professionals‖, First Edition, Wiley, 2014.

iii. Online References:

1.https://warwick.ac.uk/fac/sci/dcs/teaching/modules/cs342/
2.https://nptel.ac.in/courses/106/105/106105152/
3.https://machinelearningmastery.com/machine-learning-in-python

COURSE CODE	COURSE TITLE	L	T	P	C
2162CS140	ARTIFICIAL INTELLIGENCE	3	0	0	3

A. Preamble

Artificial Intelligence is one of the most advanced fields of computer science which involves use of Mathematics, Statistics, Information Technology and Information Sciences in discovering new information and knowledge from large databases and optimize Human effort overall. It is a new emerging interdisciplinary area of research and development which has created interest among scientists of various disciplines like Computer Science, Mathematics, Statistics, Information Technology.

B. Prerequisite Courses

Nil

C. Course Objectives

- Expound the problem solving by Searching StateSpace.
- Solve Problem by Heuristic approach.
- Solve Problem by Random and Optimal approaches
- Solving Constraint-satisfaction problem and Planning.
- Utilize Logical knowledge representation and Experiment with Uncertainty and Reasoning

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Problem solving by Searching State Space.	K2
CO2	Problem solving by Heuristic approach.	K3
CO3	Problem solving by Random and Optimal approach.	K3
CO4	Constraints oriented problem solving and planning.	K3
CO5	Dealing with Uncertainty and Reasoning	K3
CO6	Logical knowledge representation.	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	H		M	M			M	M
CO2	H		M	M			M	M
CO3	M		M	M			M	M
CO4	M		M	M			M	M
CO5	M		M	M			M	M

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction L-9 Hours

Artificial Intelligence - Introduction. State Space Search: Generate and Test- Simple Search- Depth First Search- Breadth First Search- Comparison of BFS and DFS- Quality of Solution- Depth Bounded DFS- Depth First Iterative Deepening .Heuristic Search: Heuristic Functions- Best First Search- Hill Climbing- Local Maxima- Solution Space Search- Variable Neighborhood Descent- Beam Search- Tabu Search- Peak to Peak Methods.

UNIT 2 Randomized Search And Emergent Systems L-9 Hours

Iterated Hill Climbing- Simulated Annealing- Genetic Algorithms- The Travelling Salesman Problem- Neural Networks- Emergent Systems- Ant Colony Optimization. Finding Optimal Paths: Brute Force- Branch & Bound- Refinement Search- Dijkstra's Algorithm- Algorithm A*- Admissibility of A*- Iterative Deepening A*- Recursive Best First Search- Pruning the OPEN and CLOSED List- Divide and Conquer Beam Stack Search.

UNIT 3 Planning L-9 Hours

The STRIPS Domain- Forward and Backwards State Space Planning- Goal Stack Planning- Plan Space Planning- A Unified Planning Framework. **Constraint Satisfaction Problem**: N-Queens- Constraint Propagation- Scene Labeling- Higher Order and Directional Consistency- Algorithm Backtracking- Look-ahead Strategies- Strategic Retreat.

UNIT 4 Knowledge Based Reasoning L-9 Hours

Agents- Facets of Knowledge. Logic and Inferences**:** Formal Logic- Propositional Logic- Resolution Method in Propositional Logic- First Order Logic- Incompleteness of Forward Chaining- Resolution Refutation in First Order Logic- Deductive Retrieval – Resolution Method's Complexity in FOL- Horn Clauses and SDL Resolution- Backward Chaining- Second Order Logic.

UNIT 5 Structured Knowledge Representation L-9 Hours

Hierarchies in Domain- The Scheme- Frames- Semantic Net- Scripts, Goals, Plans and MOPs- Inheritance in Taxonomies- Description Logics- Formal Concept Analysis- Conceptual Graphs.

Total: 45 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1.“A First Course in Artificial Intelligence”, by Deepak Khemani, McGraw-Hill Education, 2013.
2.Peter Norvig, and Stuart Russell, “Artificial Intelligence: A Modern Approach”, Global Edition, Pearson; 4th edition (2021)

ii. Reference Books:

1.Luke Dormehl, “Thinking Machines: The Quest for Artificial Intelligence--and Where It's Taking Us Next” 2017, TarcherPerigee

2.V S Janakiraman, “Foundations of Artificial Intelligence and Expert Systems”, 2005, Macmillan.

iii. Online References:

1.https://nptel.ac.in/courses/106/102/106102220/

2.https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-036-introduction-to-machine-learning-fall-2020/

PROGRAM SPECIFIC ELECTIVE

COURSE CODE	COURSE TITLE	L	T	P	C
2162CS201	**FOUNDATIONS OF DATA SCIENCE**	**2**	**0**	**2**	**3**

A.Preamble

To provide an in-depth understanding of high dimensional spaces and singular value decomposition

B.Prerequisite Courses

C.Course Objectives

1. To impart the fundamentals of data science.
2. To enable the students think the capability to build big-data
3. Developing design skills of models for big data problems
4. Understand and gain exposure in programming tools for data sciences

D.Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Discuss the high dimensional spaces and singular value decomposition modeling in massive business problems.	K1
CO2	Discuss the process of random walks on undirected graphs and Markov chains model.	K2
CO3	Choose appropriate machine and deep learning technique for massive data problems.	K3
CO4	Use different clustering techniques for massive data problems.	K2
CO5	Design and evaluate random graphs and branching process.	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	M		M				M	
CO2	M		M	M	M		M	
CO3	M		M	M	M		M	
CO4	M		M	M	M		M	
CO5	M		M		H			H

H- Strong; M-Medium; L-Low

F.Course Contents

UNIT 1 High-Dimensions and Singular Value Decomposition L-9 Hours

Introduction, Law of Large Numbers, Geometry of High Dimensions, Properties of the Unit Ball, Generating Points Uniformly at Random from a Ball, Gaussians in High Dimension, Random Projection and Johnson-Linden Strauss Lemma, Separating Gaussians, Fitting a Spherical Gaussian to Data, Singular Vectors, Singular Value Decomposition, Best Rank-k Approximations, Left Singular Vectors, Power Method for Singular Value Decomposition.

UNIT 2 Random Walks and Markov Chains L-9 Hours

Singular Vectors and Eigenvectors, Applications of Singular Value Decomposition, Stationary Distribution, Markov Chain Monte Carlo, Areas and Volumes, Convergence of Random Walks on Undirected Graphs, Electrical Networks and Random Walks, Random Walks on Undirected Graphs with Unit Edge Weights, Random Walks in Euclidean Space, Web as a Markov Chain.

UNIT 3 Machine Learning L-9 Hours

Overfitting and Uniform Convergence, Regularization: Penalizing Complexity, Online Learning and the Perceptron Algorithm, Kernel Functions, Online to Batch Conversion, Support-Vector Machines, VC-Dimension, Strong and Weak Learning - Boosting, Stochastic Gradient Descent, Combining (Sleeping) Expert Advice, Deep Learning, Further Current Directions.

UNIT 4 Massive Data Problems and Clustering L-9 Hours

Frequency Moments of Data Streams, Matrix Algorithms using Sampling, k-Means Clustering, k-Center Clustering, Finding Low-Error Clustering, Spectral Clustering, Approximation Stability, High-Density Clusters, Kernel Methods, Recursive Clustering based on Sparse Cuts, Dense Submatrices and Communities, Community Finding and Graph Partitioning, Spectral clustering applied to social networks

UNIT 5 Random Graphs: L-9 Hours

The G(n,p) Model, Phase Transitions, Giant Component, Cycles and Full Connectivity, Phase Transitions for Increasing Properties, Branching Processes, CNF-SAT, Non-uniform Models of Random Graphs, Growth Models, Small World Graphs.

Total: 45 Hours

A. Learning Resources (in IEEE Format)

i.Text Books:

1. Avrim Blum, John Hopcroft, and Ravindran Kannan, "Foundations of Data Science", Draft Version, June, 2017.
2. Mining of Massive Datasets. v2.1, Jure Leskovek, Anand Rajaraman and JefreyUllman, Cambridge University Press. (2019)

ii.Reference Books:

1. Jonathan Dinu, "Foundations of Data Science: A Practical Introduction to Data Science with Python", Addison Wesley (2 July 2019), ISBN-10: 0134398807.

iii.Online Resources:

1. https://web.stanford.edu/class/msande226/l_notes.html
2. https://www.cl.cam.ac.uk/teaching/1819/DataSci/materials.html

COURSE CODE	COURSE TITLE	L	T	P	C
2162CS202	BIG DATA ANALYTICS	2	0	2	3

A.Preamble

Understand how to leverage the insights from big data analytics. Analyze data by utilizing various statistical and data mining approaches and perform analytics on real-time streaming data. Understand the various NoSql alternative database models.

B.Prerequisite Courses

C.Course Objectives

- Understand the Big Data Platform and its Use cases
- Provide an overview of Apache Hadoop
- Provide HDFS Concepts and Interfacing with HDFS
- Understand Map Reduce Jobs
- Provide hands on Hodoop Eco System

D.Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	To understand the competitive advantages of big data analytics	K2
CO2	To understand the big data frameworks	K3
CO3	To learn data analysis methods	K3
CO4	To learn about stream computing	K3
CO5	To gain knowledge on Hadoop related tools such as HBase, Cassandra, Pig, and Hive for big data analytics	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	H		M	M			M	M
CO2	H		M	M			M	M
CO3	M		M	M			M	M
CO4	M		M	M			M	M
CO5	M		M	M			M	M

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction To Big Data L-9 Hours

Big Data – Definition, Characteristic Features – Big Data Applications - Big Data vs Traditional Data - Risks of Big Data - Structure of Big Data - Challenges of Conventional Systems – Web Data – Evolution of Analytic Scalability - Evolution of Analytic Processes, Tools and methods -Analysis vs Reporting - Modern Data Analytic Tools.

UNIT 2 Hadoop Framework L-9 Hours

Distributed File Systems - Large-Scale File System Organization – HDFS concepts – MapReduce Execution, Algorithms using MapReduce, Matrix-Vector Multiplication – Hadoop YARN

UNIT 3 Data Analysis L-9 Hours

Statistical Methods: Regression modelling, Multivariate Analysis - Classification: - Rule Mining - Cluster Analysis, Types of Data in Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density Based Methods, Grid Based Methods, Model Based Clustering Methods, Clustering High Dimensional Data - Predictive Analytics – Data analysis using R.

UNIT 4 Mining Data Streams L-9 Hours

Streams: Concepts – Stream Data Model and Architecture - Sampling data in a stream – Mining Data Streams and Mining Time-series data - Real Time Analytics Platform (RTAP) Applications -Case Studies - Real Time Sentiment Analysis, Stock Market Predictions.

UNIT 5 Big Data Tools L-9 Hours

Introduction to NoSQL – Aggregate Data Models – HBase: Data Model and Implementations – HBase Clients – Examples –. Cassandra: Data Model – Examples – Cassandra Clients – Hadoop Integration. Pig – Grunt – Pig Data Model – Pig Latin – developing and testing Pig Latin scripts. Hive – Data Types and File Formats – HiveQL Data Definition – HiveQL Data Manipulation – HiveQL Queries

Total: 45 Hours

A. Learning Resources (in IEEE Format)

i. Text Books:

1.“Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, Second Edition, 2014.
2.Michael Minelli, Michelle Chambers, and AmbigaDhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley, 2013.

ii. Reference Books:

1.Bill Franks,Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics‖, Wiley and SAS Business Series, 2012.
2,David Loshin, Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph, 2013.
3.Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, Second Edition, 2007.

iii.Online References:

1.https://stat.mit.edu/academics/data-science-data-insight-professional-education/
2.https://nptel.ac.in/courses/106/104/106104189/

COURSE CODE	COURSE TITLE	L	T	P	C
2162CS203	**DATA VISUALIZATION**	**2**	**0**	**2**	**3**

A.Preamble

Explain principles of visual perception and to apply core skills for visual analysis. Apply appropriate visualization techniques for various data analysis tasks. To design information dashboard

B.Prerequisite Courses

NIL

C.Course Objectives

- Know the basics of data visualization
- Understand the importance of data visualization and the design and use of many visual components
- Learn to wisely use various visualization structures such as tables, spatial data, time-varying data, tree and network, etc.
- Learn the basics of colors, views, and other popular and important visualization-based issues.
- Learn basic algorithms in data visualization

D.Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	To develop skills to both design and critique visualizations	K2
CO2	To introduce visual perception and core skills for visual analysis	K2
CO3	To understand visualization for time-series, ranking and deviation analysis.	K3
CO4	To understand visualization for distribution, correlation and multivariate analysis.	K3
CO5	To understand issues and best practices in information dashboard design	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	**H**		**M**	**M**			**M**	**M**
CO2	**H**		**M**	**M**	**M**		**M**	**M**
CO3	**H**		**M**	**M**	**M**		**M**	**M**
CO4	**H**		**M**	**M**	**M**		**M**	**M**
CO5	**H**		**M**	**M**	**M**		**M**	**M**

H- Strong; M-Medium; L-Low

F.Course Contents

UNIT 1 Introduction L-9Hours

Context of data visualization – Definition, Methodology, Visualization design objectives. Key Factors – Purpose, visualization function and tone, visualization design options – Data representation, Data Presentation, Seven stages of data visualization, widgets, data visualization tools.

UNIT 2 Visualizing Data Methods L-9Hours

Mapping - Time series - Connections and correlations - Scatterplot maps - Trees, Hierarchies and Recursion - Networks and Graphs, Info graphics

UNIT 3 Time-Series, Ranking, And Deviation Analysis L-9Hours

Time-series analysis – time-series patterns – time-series displays – time-series best practices – part-to-whole and ranking patterns – part-to-whole and ranking displays – best practices – deviation analysis – deviation analysis displays – deviation analysis best practices.

UNIT 4 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 – multivariate analysis – multivariate patterns – multivariate displays – multivariate analysis techniques and best practices.

UNIT 5 Information Dashboard Design L-9 Hours

Information dashboard – Introduction– dashboard design issues and assessment of needs – Considerations for designing dashboard-visual perception – Achieving eloquence.

.

Total: 45 Hours

G.Learning Resources (in IEEE Format)

i. Text Books:

1. Stephen Few, now you see it: Simple Visualization techniques for quantitative analysis, Analytics Press, 2009.
2. 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, Second Edition, Graphics Press, 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
2162CS205	**DEEP LEARNING**	**2**	**0**	**2**	**3**

A. Preamble

This course covers the basics of neural networks are included. Artificial Neural Network, Convolutional Neural Network, Recurrent Neural Networks are included Deep learning tool and its application are included.

B. Prerequisite Courses

2161CS129 Machine Learning Techniques

C. Course Objectives

- To introduce the fundamentals of neural networks
- To enable the students to understand Recurrent Neural networks and Convolutional Neural Networks.
- To introduce complex learning models and deep learning models
- Explore different software packages and tools for Deep learning models

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Analyze Artificial Neural Network model and implement the perceptron learning algorithms	K3
CO2	Apply the Convolutional Neural Network Model to build the deep learning applications	K3
CO3	Implement Recurrent Neural Network and LSTM model and its applications.	K3
CO4	Understand the concept of deep Neural Networks and GAN Architecture	K2
CO5	Explore deep learning tools and applications	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	PO6	PSO1	PSO2
CO1	**H**		**M**	**M**			**M**	**M**
CO2	**H**		**M**	**M**			**M**	**M**
CO3	**H**		**M**	**M**		**M**	**M**	**M**
CO4	**H**		**M**	**M**			**M**	**M**
CO5	**H**		**M**	**M**	**M**		**M**	**M**

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Artificial Neural Networks L-9 Hours

Artificial Neural Networks: Introduction, Single and Multi-layer Perceptron, Perceptron Training Rule, Activation Functions and softmax cross entropy loss function - Gradient Descent, Backpropagation, Radial basis Function Network

UNIT 2 Convolutional Neural Networks Architecture L-9Hours

Introduction to Convolutional Neural Networks, Principles behind CNNs, Multiple Filters, Kernel filter, Convolution and Pooling Operation, CNN applications. CNN Architecture- Alex Net, ZFNet, VGG, C3D, Google Net, ResNet, Mobile Net

UNIT 3 Recurrent Neural Networks L-9Hours

Introduction to Recurrent Neural Networks: Introduction to RNNs, Unfolded RNNs,Seq2Seq RNNs, LSTM, RNN applications – Word Embedding

UNIT 4 Deep Neural Network L-9Hours

Improving Deep NN – Hyperparameter Tuning, Regularization and Optimization, Types of errors, Bias Variance Trade-off, Overfitting and Underfitting.

UNIT 5 Deep Learning Tools And Applications L-9 Hours

Introduction to TensorFlow: Computational Graph, Key highlights, Creating a Graph, Regression example, Gradient Descent, Tensor Board, Modularity, Sharing Variables, Kera's .Deep Learning applications: Image Processing, Natural Language Processing, Speech
Recognition, Video Analytics.

Total: 45 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1. Ian Goodfellow, YoshuaBengio, Aaron Courville, Deep Learning, MIT Press, 2016.
2. CosmaRohillaShalizi, Advanced Data Analysis from an Elementary Point of View, 2015.

ii. Reference Books:

1. Bishop, C.,M., Pattern Recognition and Machine Learning, Springer, 2006.
2. Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.
3. Golub, G.,H., and Van Loan,C.,F., Matrix Computations, JHU Press,2013.
4. Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education,2004

iii. Online References:

1. https://nptel.ac.in/courses/106/106/106106184/
2. https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-s191-introduction-to-deep-learning-january-iap-2020/

COURSE CODE	COURSE TITLE	L	T	P	C
2162CS210	Health Care Analytics	2	0	2	3

A. Preamble

The course discusses the current and future applications of data science in healthcare with the goal of learning to bring computer technologies into the clinic safely and ethically.

B. Prerequisite Courses

Big data Analytics

C. Course Objectives

Students are able to

Understand the theories and practices adopted in healthcare Information Systems in the light of medical standards, medical data formats and recent trends in Hospital Information Systems.

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Understand about health informatics and different ICT applications in medicine.	K2
CO2	Apply various medical standards and metrics in healthcare applications	K3
CO3	Analyze the function of Hospital Information Systems.	K3
CO4	Evaluate the healthcare informatics tools and techniques.	K3
CO5	Analyze the concept and need of different healthcare information systems.	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	H		M	M			M	
CO2	H		M	M			M	
CO3	H		M	M	M		M	
CO4	H		M	M			M	
CO5	H		M	M			M	

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction to Health Care Informatics L-9 Hours

Introduction - Health Informatics - Structure of Medical Informatics –Functional capabilities of Hospital Information System - On-line services and off – line services - History taken by computer, Dialogue with the computer.

UNIT 2 Health Data Standards and Metrics L-9Hours

Evolution of Medical Standards – IEEE 11073 - HL7 – DICOM – IRMA - LOINC – HIPPA – Electronics Patient Records –Healthcare Standard Organizations – JCAHO (Join Commission on Accreditation of Healthcare Organization) - JCIA (Joint Commission International Accreditation) - Evidence Based Medicine - Bioethics.

UNIT 3 Healthcare Data Acquisition and Storage L-9 Hours

Plug-in Data Acquisition and Control Boards – Data Acquisition using Serial Interface - Medical Data formats – Signal, Image and Video Formats – Medical Databases - Automation in clinical laboratories - Intelligent Laboratory Information System - PACS , Data mining.

UNIT 4 Healthcare Information Retrieval L-9 Hours

Bioinformatics Databases, Bio-information technologies, Semantic web and Bioinformatics, Genome projects, Clinical informatics, Nursing informatics, Public health informatics -Education and Training.

UNIT 5 Recent Trends in Healthcare Informatics L-9 Hours

Medical Expert Systems, Virtual reality applications in medicine, Virtual Environment - Surgical simulation - Radiation therapy and planning – Telemedicine – virtual Hospitals - Smart Medical Homes – Personalized e-health services – Biometrics - GRID and Cloud Computing in Medicine.

TOTAL: 45 Hours

G. Laboratory Experiments

Laboratory Experiments TOTAL:30 Hours

Part –I:

Task 1: Create a program to display grayscale and color medical image using read and write operation.

Task 2: Create a vision program to find histogram value and display histograph of a grayscale and color medical image.

Task 3: Create a vision program for Non-Linear Filtering technique using edge detection on medical image

Task 4: Create a program to discretize a medical image using Fourier transform and discrete wavelet transform.
Task -5: Disease recognition using simple machine learning models
Task 6: Create a program for segmentation of a medical image using watershed transforms.
Task 7: Diagnosis of diseases in scanned images using Image segmentation techniques.
Task 8: Apply deep supervised learning techniques to diagnose disorders compare their performance
Task 9: Build a risk assessment model using linear and tree-based models
Task 10: Develop chat bot using natural language techniques

Part –II:

Task 1: Write a program to display grayscale and color Lung PET/CTmedical image using read and write operation.
Task 2: Write a program to find histogram value and display histograph of a grayscale and color CT Colongraphy medical image.
Task 3: Write a program for Non-Linear Filtering technique using edge detection on Virtual Colonoscopy medical image
Task 4: Write a program to discretize a Virtual Colonoscopy medical image using Fourier transform and discrete wavelet transform.
Task -5: Disease recognition using simple machine learning models on PET/CT phantom scan collection dataset
Task 6: Create a program for segmentation of a Lung Image Database Consortium (LIDC) medical image using watershed transforms.
Task 7: Diagnosis of diseases in scanned Breast MRI images using Image segmentation techniques.
Task 8: Apply deep supervised learning techniques to diagnose disorders Lung Image Database Consortium (LIDC) dataset and compare their performance
Task 9: Build a risk assessment model using linear and tree-based models heart disease dataset
Task 10: Develop chat bot using natural language techniques for hospital assistant.

TOTAL: 75 Hours

H. Learning Resources (in IEEE Format)

i. Text Books:

1. R.D.Lele, “Computers in Medicine: Progress in Medical Informatics’’, Tata McGraw Hill Publishing computers Ltd, New Delhi, 2017.

ii. Reference Books:

1. Mohan Bansal, “Medical informatics”, Tata McGraw Hill Publishing computers Ltd, New Delhi, 2003.
2. N.Mathivanan, “PC-Based Instrumentation’’, Prentice Hall of India Pvt Ltd – New Delhi, 2007.
3. Wager, K. A., Lee, F. W., & Glaser, J. P, Health care information Systems: A practical approach for health care management -4th Edition, Jossey-Bass, 2017

iii. Online References:

1. “Coursera course on Fundamentals of Machine Learning for Healthcare”, Accessed on Apr. 20, 2021 [Online], https://www.coursera.org/learn/ fundamental-machine-learning-healthcare
2. “Coursera course on Evaluations of AI Applications in Healthcare”, Accessed on Apr. 20, 2021 [Online], https://www.coursera.org/learn/evaluations-ai-applications-healthcare

COURSE CODE	COURSE TITLE	L	T	P	C
2162CS211	TIME SERIES AND FORECASTING	2	0	2	3

A. Preamble

The 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

10211CS110 – Data Warehousing and Data mining

C. Course Objectives

- To provide students with a framework that will help them choose the appropriate descriptive statisticsin various data analysis situations.
- To analyze distributions and relationships of real-time data.
- To apply estimation and testing methods to make inference and modeling techniques for decisionmaking using various techniques including multivariate analysis.

D. Course Outcomes

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

CO No's	Course Outcomes	K - Level
CO1	Compute the elimination techniques to remove trend and seasonality in timeseries analysis.	K3
CO2	Apply exploratory techniques, do simulation on time series data and perform how to store temporal data.	K3
CO3	Use the various time series models to solve problems.	K3
CO4	Implement the Machine Learning and Deep Learning techniques on timeseries data	K3
CO5	Calculate suitable metrics 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	H		M	M			M	M
CO2	H		M	M			M	M
CO3	M		H	M			M	M
CO4	H		M	M	M		M	M
CO5	H		M	M			M	M

H- Strong; M-Medium; L-Low

F. Course Contents

UNIT 1 Introduction To Time Series Analysis L-9Hours

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

UNIT2 Exploratory Data Analysis, Simulation, And Storing L-9Hours

Augmented Dickey–Fuller test - Rolling windows – Autocorrelation - Partial Autocorrelation – Spurious Correlations – Simulating Time Series Data: Simulation Versus Forecasting - Monte Carlo Simulations

UNIT 3 Stationary And Non Stationary Time Series Model L-9Hours

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

UNIT 4 Machine Learning And Deep 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 - 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 - Deep Learning for Time Series: CNN, RNN.

UNIT 5 Performance Analysis L-6 Hours

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 and Case studies

TOTAL:30 Hours

G. Learning Resources (in IEEE Format)

i. Text Books:

1.Aileen Nielsen, “Practical Time Series Analysis - Prediction with Statistics and Machine Learning”, O’Reilly publications, First Edition, 2019. (UNIT II, IV, V)
2. Peter J. Brockwell, Richard A. Davis, “Introduction to Time Series and Forecasting”, Second Edition, Springer, 2016 (UNIT I, III**Reference Books:**

ii. References:

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
3. James D. Hamilton, “Time Series Analysis”, Princeton University Press, First Edition, 1994

iii. Online Resource(s) :

1. https://nptel.ac.in/noc/courses/noc17/SEM1/noc17-ch03/
2. https://ocw.mit.edu/courses/economics/14-384-time-series-analysis-fall-2013/lecture-notes/