UNIVERSITY SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PROGRAMME STRUCTURE

M.TECH. COMPUTER SCIENCE & ENGINEERING

SPECIALIZATION: **DATA SCIENCE**

2022-2024

GAUTAM BUDDHA UNIVERSITY

GAUTAM BUDH NAGAR, GREATER NOIDA, UP, INDIA

Semester 1

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CS521	ADVANCE DATABASE MANAGEMENT SYSTEM	3	0	0	3	CC1
2	CS523	DESIGN AND ARCHITECTURE OF SOFTWARE SYSTEM	3	0	0	3	CC2 / FC
3	CS525	ADVANCED DATA STRUCTURE AND ALGORITHM	3	1	0	4	CC3
4	CS527	RESEARCH TECHNIQUES IN ICT	3	0	0	3	CC4
5	CS529	JAVA PROGRAMMING	3	0	0	3	CC5
6	ES415	ENERGY AND ENVIRONMENT	3	0	0	3	OE1 /AECC
7	EN531	Language, Culture and Society	3	0	0	3	OE2 /AECC
8	CS581	ADVANCED DATABASE MANAGEMENT SYSTEM LAB	0	0	3	2	CC-L1
9	CS583	Java Programming Lab	0	0	3	2	CC-L2 / SEC
10	GP	General Proficiency	Non-Credit				
Total Hours and Credits			21	1	6	26	

Semester 2

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD522	PYTHON PROGRAMMING	3	0	0	3	CC6
2	CD524	AI ENABLED DATA SCIENCE	3	0	0	3	CC7
3	CD526	OPTIMIZATION TECHNIQUES FOR DATA SCIENCE	3	0	0	3	CC8
4	CD528	ADVANCED MS EXCEL	3	0	0	3	CC9
5		Elective-1	3	0	0	3	E1 / DSE
6		Elective-2	3	0	0	3	E2 / DSE
7		Generic Elective	3	1	0	4	GE1
8	CD584	ADVANCED MS EXCEL LAB	0	0	3	2	CC-L3 / SEC
9	CD582	PYTHON PROGRAMMING LAB	0	0	3	2	CC-L4
10	GP	General Proficiency	Non-Credit				
Total Hours and Credits			21	1	6	26	

S.No	Course Code	Course Name	L	T	P	Credi	Types
1	CD621	MACHINE LEARNING ALGORITHMS	3	0	0	3	CC10
2	CD623	BIG DATA FRAMEWORK FOR DATA SCIENCE	3	0	0	3	CC11
3		Elective-3	3	0	0	3	E3 / DSE
4		Elective-4	3	0	0	3	E4 / DSE
5	CD681	Big Data Framework for Data Science Lab	0	0	3	2	CC-L5
6	CA683	Summer Project	0	0	8	4	SP / E
7	CA691	Dissertation Part - I	0	0	16	8	DP1 / E
8	GP	General Proficiency	Non-Credit				
Total Hours and Credits			12	0	27	26	

Semester 3

Semester-4

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD692	Dissertation Part - II	0	0	52	26	DP2 / E
2	GP	General Proficiency	Non-Credit				
Total Hours and Credits			0	0	52	26	

GRAND TOTAL OF CREDITS = 104

ELECTIVES FROM DCSE

S.No.	Course Code	Course Name	L	T	P	Credits	Types
1	CD530	Modeling Process in Data Science	3	0	0	3	E1
2	CD532	Data Security	3	0	0	3	E1
3	CD534	Parallel and Distributed Computing	3	0	0	3	E1
4	CD536	Data Stream Management	3	0	0	3	E1
5	CD538	Predictive Analytics and Data Visualization	3	0	0	3	E2
6	CD540	Business Acumen and Intelligence	3	0	0	3	E2
7	CD542	Data Science for Business Analytics	3	0	0	3	E2
8	CD544	Advanced Semantics Web Technology	3	0	0	3	E2
9	CS625	Big Data Analytics	3	0	0	3	E3
10	CS627	Advanced Metrics for Economics	3	0	0	3	E3 E3
11	CS629	Multivariate Techniques for Data Science	3	0	0	3	E3
12	CS631	Block chain Technology for Data Science	3	0	0	3	E4
13	CS633	Deep Learning in Biomedical and Genomics	3	0	0	3	E4
14	CS635	Data Analytics for Healthcare	3	0	0	3	E4
15	CS637	Regression for Time Series Analysis	3	0	0	3	E4
16	CS639	Internet of Things	3	0	0	3	E4

OPEN AND GENERIC ELECTIVES FROM OTHER SCHOOLS							
1	ES415	Energy and Environment	3	0	0	3	OE1
2	EN531	Language, Culture and Society	3	0	0	3	OE2
3	MA402	Modeling and Simulation	3	1	0	4	GE1
4	MA416	Probability and Stochastic Process	3	1	0	4	GE1

CA Computer Science and Artificial Intelligence for Course Code

CS Computer Science for Course Code

CC Core Course from USICT for Type of Course

CC-L Core Course Lab from USICT for Type of Course

GE General Elective from related discipline of other Deptt. /School

AECC Ability Enhancement Compulsory Course

OE Open Elective from other discipline of other Deptt. /School

FC Foundation Course

DSC Discipline Specific Course

SEC Skill Enhancement Course

E Elective

DP1 Dissertation Part 1

DP2 Dissertation Part 2

SP Summer Project

Semester-I

ADVANCE DATABASE MANAGEMENT SYSTEM			
Course Code:	**CS521**	**Course Credits:**	**3**
Course Category:	**CC1**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1 Knowledge of database design
2 A general understanding of database ,design and dependency
3 Understanding of different types of databases
4 Knowledge of databases on the internet
5 Application on enhanced database

COURSE OUTCOMES
At the end of the course the students should be able to:
1 Basic knowledge and understanding of ER diagram and UML class diagram.
2 Ability to apply functionality and Normalization on relational database.
3 Understand and fetch data from object oriented , parallel and distributed databases.
4 Use XML and understand unstructured data
5 Implement concept and deduction of enhanced database on different applications

UNIT I INTRODUCTION TO DATABASE DESIGN

Entities, Attributes, Entity Sets, Relationships, Key Constraints, Participation Constraints, Weak Entities, UML Class Diagrams, Subclasses, Superclasses, Inheritance,Specialization, Generalization, Constraints and Characteristics of Specialization and Generalization Hierarchies, Modeling of UNION Types Using Categories, Representing Specialization and Generalization In UML Class Diagrams, Data Abstraction, Knowledge Representation and Ontology Concepts.

UNIT II DATABASES DESIGN THEORY

Problems Caused by Redundancy, Decompositions, Problems Related to Decomposition, Reasoning About FD's, FIRST, SECOND, THIRD Normal Form, BCNF, Forth Normal Form, Lossless Join Decomposition, Dependency Preserving Decomposition, Schema Refinement in DataBase Design, Multi Valued Dependencies.

UNIT III OBJECT- ORIENTED, PARALLEL AND DISTRIBUTED DATABASES

Overview of Object-Oriented Concepts, Object Identity, Object Structure, Type Constructor, Encapsulation of Operations, Methods and Persistence; Architectures For Parallel Databases, Parallel Query Evaluation, Parallelizing Individual Operations, Sorting Joins, Distributed Database Concepts,

Data Fragmentation, Replication and Allocation Techniques for Distributed Database Design, Query Processing in Distributed Databases, Concurrency Control and Recovery in Distributed Databases.

UNIT IV DATABASES ON THE WEB AND SEMI–STRUCTURED DATA Web interface, XML, structure of XML data, querying XML data, storage of XML data, XML applications, semi-structured data model, indexes for text data.

UNIT V ENHANCED DATA MODELS FOR ADVANCED APPLICATIONS

Active database concepts, temporal database concepts, spatial databases: concept and architecture, deductive databases and query processing, mobile databases, Geographic Information Systems (GIS).

Text Books:

1. Elmasri and Navathe, Fundamentals of Database Systems,
2. Ramakrishnan and Gehrke, Database Management Systems,

References Books:

3. Korth, Silberschatz, Sudarshan, Database System Concepts,
4. Rob and Coronel, Database Systems: Design, Implementation and Management,
5. Date and Longman, Introduction to Database Systems,

DESIGN AND ARCHITECTURE OF SOFTWARE SYSTEM			
Course Code:	**CS-523**	**Course Credits:**	**3**
Course Category:	**CC**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1 Understand the creational and structural patterns.
2 Be capable of applying his knowledge to create an architecture for given application.
3 Be able to explain the role of analyzing architectures.
4 Understanding of software design.
5 Be able to identify different structural patterns.

COURSE OUTCOMES
At the end of the course the students should be able to:
1 Understand the architecture, creating it and moving from one to any, different structural patterns.
2 Ability to analyze the architecture and build the system from the components.
3 Design creational and structural patterns.
4 Learn about behavioral patterns.
5 Do a case study in utilizing architectural structures.

UNIT I SOFTWARE ARCHITECTURE

Foundations of software architecture, goals of software architecture limitations, role of software architect, types of architecture, qualities attributes, qualities scenario, architectural styles, common architectural design, architectural design process, key architecture principles, key design principles, functional and non-functional properties of software architectures, heterogeneous architectures, virtual machine architecture, data flow architecture, service- oriented architecture.

UNIT II DESIGN FUNDAMENTALS AND METHODOLOGIES

Nature of design process: objectives, building modules, constructs, design qualities, assessing the design, design viewpoints for software, design strategies: top down and bottom up, organizational methods and design, Jackson structural programming, Jackson system development, models for software architecture

UNIT III SOFTWARE ARCHITECTURE DESIGN

Architectural design and mapping, architecture design patterns, module architecture view, styles of the module view type, execution architecture view, code architecture view

mponent-and-connector view type, styles of component-and-connector view type, allocation view type and styles, object-oriented architecture, user interface architecture, quantified design space, formalizing architectural description language, first class connectors, tools for architectural design: Unicon, A4;

exploiting style in architectural design, architectural interconnection.

UNIT IV INTERACTION ORIENTED SOFTWARE ARCHITECTURE AND DESIGN

Model-View-Controller (MVC), Presentation-Abstraction-Control (PAC) architecture, distributed architecture: client server architecture, multi-tire, service-oriented architecture (SOA). Design principles, traditional approach to design, Structured Analysis Design Technique (SADT), Structures System Analysis and Design Method (SSADM), user interface design; human factor, human computer interaction, interface design guide lines, standards, object-oriented analysis and design.

UNIT V PATTERNS

Design patterns, creational patterns, access control patterns, service variation patterns, service extension patterns, archetypes patterns, model driven architecture with archetype patterns, literate modeling, Customer Relationship Management (CRM) archetype pattern, product archetype pattern, quantity archetype pattern, rule archetype pattern, layering, organizing domain logic, mapping to relational databases, web presentation, domain logic patterns, data source architectural patterns, object-relational behavioral patterns, object relational structural patterns, object-relational metadata mapping patterns, web presentation patterns, distribution patterns, offline concurrency patterns.

Text Books:

1. Software Architecture Perspectives on an Emerging Discipline, M. Shaw Prentice-Hall, 1996.
2. Software Architecture Design: Methodology and Styles, Lixin Tao, Xiang Fu and Kai Qian, Stipes Publishing L.L.C., 2006.
3. Software Architecture in Practice, Len Bass, Paul Clements, Rick Kazman, Pearson Education Asia,2003.

References Books:

4. Software Design, David Budgen, Addison-Wesley, 1994.
5. Software Engineering, Pressman R.S, McGraw Hill Inc., 1996.

ADVANCED DATA STRUCTURE AND ALGORITHM			
Course Code:	**CS525**	**Course Credits:**	**4**
Course Category:	**CC3**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 01**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 15**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1 Understand the appropriate data structures, ADT libraries, and use it to design algorithms for a specific problem.
2 Be capable of solving problems using abstraction techniques.
3 Be able to choose appropriate algorithms for a specific problem.
4 Be able to analyze algorithms in terms of their efficiency and correctness.
5 Understanding the recent developments in the area of algorithm design.

COURSE OUTCOMES
At the end of the course the students should be able to:
1 Design and analyze programming problem statements.
2 Choose appropriate data structures and algorithms for a specific problem.
3 Understand the necessary mathematical abstraction to solve problems.
4 Come up with analysis of efficiency and proofs of correctness.
5 Comprehend and select algorithm design approaches in a problem specific manner.

UNIT I INTRODUCTION

Review of Basic Concepts: Abstract data types, Data structures, Algorithms, Big-Oh, Small-Oh, Omega, Small-Omega and Theta Notations, finding time complexity of programs, **Recurrence Relations:** Solving Recurrence Relations, Substitution Method, Master Theorem.

UNIT II Hashing

Review of Hashing, Hash Function, Collision Resolution Techniques in Hashing, Separate Chaining, Open Addressing, Linear Probing, Quadratic Probing, Double Hashing, Rehashing, Extendible Hashing, Recent Trends in Hashing.

UNIT III TREES & GRAPH

Trees: Binary Search Trees, AVL Trees, Red Black Trees, 2-3 Trees, B-Trees, Splay Trees, Minimum Spanning Tree (MST), Kruskal's Algorithm and Prim's Algorithm, Applications to MST.

Graph: Graph, Breadth First Search, Depth First Search, Shortest path in edge-weighted case (Dijkstra's), Bellman Ford Algorithms, Topological Sorting.

UNIT IV SELECTED TOPICS

Strassen's Matrix Multiplication, Greedy method VS Dynamic Programming, Job sequencing with deadlines, Fractional Knapsack Problem, 0/1 Knapsack Problem, Travelling Salesman Problem, Huffman coding, Pre order, Post order, Inorder traversal, Postfix to infix notation, Infix to Postfix notation.

UNIT V Linear Programming

Geometry of the feasibility region and Simplex algorithm

NP-completeness: Examples, proof of NP-hardness and NP-completeness.

Recent Trends: Recent Trends in problem solving paradigms using recent searching and sorting techniques by applying recently proposed data structures.

Text Books:

1. Introduction to Algorithms, Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein.
2. Algorithms Unlocked: Thomas H. Cormen.
3. The Algorithm Design Manual, Steven S. Skiena.

References Books:

4. Algorithms: Robert Sedgewick and Kevin Wayne.
5. Advanced Data Structures: Peter Brass.

RESEARCH TECHNIQUES IN ICT			
Course Code:	**CS527**	**Course Credits:**	**3**
Course Category:	**CC4**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1 To understand what research is.
2 To understand research methodology.
3 To understand the proper steps to be followed for report and paper writing.
4 To understand different models and algorithms in research.
5 To understand the various simulation and soft computing techniques.

COURSE OUTCOMES
At the end of the course the students should be able to:
1 Able to clearly define the research and its importance.
2 Able to apply research techniques in various domains.
3 Able to enhance the standards of society by finding solutions for a particular problem.
4 Able to write research papers so as to develop a great career in the field of research.
5 Able to understand different simulation packages.

UNIT I INTRODUCTION TO RESEARCH TECHNIQUES

Meaning of research, objectives of research, motivation in research, types of research, characteristics and prerequisites of research, significance of research, research process, sources of research problem, criteria of identifying the problem, necessity of defining the problem, errors in selecting research problem, technique involved in defining the problem, report and paper writing.

UNIT II DATA ANALYSIS ANDSTATISTICAL TECHNIQUES

Data and their analyses, quantitative methods and techniques, Measure of central tendency, measures of variation, frequency distribution, analysis of variance, methods, Correlation analysis, regression analysis, time series and forecasting, introduction to discriminant analysis, factor analysis, cluster analysis, conjoint analysis, probability distribution, binomial distribution, poisson distribution, uniform distribution, exponential distribution, and normal distribution, sampling methods, test of hypothesis.

UNIT III MATHEMATICAM MODELING

Steps of modeling, operations research models like queuing theory, stochastic processes, application of models, conceptual framework development and validation techniques, optimizationtechniques.

UNIT IV ALGORITHMIC RESEARCH

Algorithmic research problems, types of algorithmic research, types of solution procedure, steps of development of algorithm, steps of algorithmic research, design of experiments.

UNIT V SIMULATION AND SOFT COMPUTING TECHNIQUES

Introduction to soft computing, artificial neural network, genetic algorithm, fuzzy logic and their applications, tools of soft computing, need for simulation, types of simulation, simulation language, fitting the problem to simulation study, simulation models, output analysis, data simulation packages like MATLAB, NS2, ANSYS, Cadence.

Text books:

1. Research Methodology: Methods and Techniques, C.R. Kothari

Reference Books:

1. Research Methodologies, R. Panneerselvam, Prentice Hall, 2007.
2. Research in Education, Best John V. and James V Kahn, Wiley eastern, 2005.
3. Elements of Educational Research, Sukhia, S.P., P.V. Mehrotra, and R.N. Mehrotra, PHI publication, 2003.
4. Methodology of Research Education, K. Setia, IEEE publication, 2004.
5. Research methodology, Methods and Techniques, Kothari, C.R., 2000.

JAVA PROGRAMMING			
Course Code:	**CS529**	**Course Credits:**	**3**
Course Category:	**CC5**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1 Knowledge of basic Object-Oriented paradigm, practices and application.
2 A general understanding of class, object and methods.
3 Understanding of multithreading and applet.
4 Basic knowledge of swings and Beans with implementation.
5 Understanding of Servlet programming.

COURSE OUTCOMES
At the end of the course the students should be able to:
1 Basic knowledge and understanding of object-oriented programming.
2 Ability to apply OOPs concept in real life problem.
3 Ability to design, develop, maintain and evaluate large-scale software systems.
4 To produce efficient, reliable, robust and cost-effective software solutions using Java.
5 Ability to perform independent research and analysis.

UNIT I OBJECT-ORIENTED PROGRAMMING

Concept of object-oriented programming (OOP), benefits of OOP, application of OOP, Java history, Java features, Data types, Java key words, identifiers, constants, variables, declaration and scope of the variable, symbolic constant, type casting, arrays, strings, vectors, wrappers classes, operator, expressions, program control statements: decision making and branching: if, if….else, else….if, else if ladder, switch, decision making and looping: while, do….while, for.

UNIT II CLASSES, OBJECTS AND METHODS

Java class libraries, class fundamentals, object, methods, adding variables, add methods, creating objects, accessing class members, constructors, methods overloading, static members, nesting of methods, inheritance: extending a class, overriding methods, final variables and methods, final classes, finalizer methods, abstract methods and classes, visibility control, exception handling fundamental, Interface and Packages.

UNIT III MULTITHREADING AND APPLET PROGRAMMING

Multithreading programming: creating threads, thread class and runnable interface extending the thread class, stopping and blocking a thread, life cycle of a thread, thread

exceptions, thread priority, synchronization, thread communication using notify(), wait(), and notify all(), applet programming : applet basic, applets architecture, a complete applet skeleton, building applets code, applets life cycle, creating a executable applet, designing a web page, applets tag, passing parameters to applets, applets and HTML.

UNIT IV SWING AND BEANS

Introduction to Swing, Differences between AWT Controls & Swing Controls, JApplet, Swing Button: JButton, JToggleButton, CheckBoxes, Radio Button, JComboBox, Text Boxes etc., Icons, Labels, JTabbed Pains, JScroll Pains, JList, JTrees, JTables Java Beans: Introduction to Java Beans, Advantages of Java Beans, BDK Introspection, Developing a Home page using Applet & Swing.

UNIT V SERVLET PROGRAMMING

Introduction to Servlets: Lifecycle of a Servlet, The Servlet API, The javax. Servlet Package, Reading Servelet parameters, Reading Initialization parameters; The javax.servlet HTTP package, Handling Http Request & Responses, Security Issues Introduction to JSP, Problem with Servlet. The Anatomy of a JSP Page, JSP Processing. JSP Application Design with MVC Setting Up and JSP Environment: Installing the Java Software Development Kit, Tomcat Server & Testing Tomcat. U

REFERENCE BOOKS:

1. Programming with JAVA, E. Balagurusawamy, Tata McGraw Hill, 1998.
2. JAVA Beginner's guide, Herbert Schildt, Tata McGraw Hill, 2007.
3. Java How to Program, Deitel & Deitel, Prentice-Hall, 1999.
4. The Complete Reference JAVA 2, Herbert Schildt, 5th Edition, Tata McGraw Hill, 2002.
5. The Complete Reference JAVA 2, Herbert Schildt, 7th Edition, Tata McGraw Hill, 2009.
6. The Java Programming Language, Ken Arnold, James Gosling, Addison-Wesley, 1996.
7. How to Program Java, Peter Coffee, Ziff-Davis Press, 1996.

ADVANCED DATABASE MANAGEMENT SYSTEM LAB			
Course Code:	**CS581**	**Course Credits:**	**2**
Course Category:	**CC-L1**	**Course (U / P)**	**U / P**
Course Year (U / P):	**4U / 1P**	**Course Semester (U / P):**	**8U / 2P**
No. of Lab (Hrs/Week) / Total No. of Lab	**03 / 10**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. To explore the features of a Database Management Systems
2. To interface a database with front end tools
3. To understand the internals of a database system
4. To provide a strong foundation in advanced database concepts from an industry perspective.
To learn query processing and transaction management concepts for object-relational database and distributed database

COURSE OUTCOME
At the end of the course the students should be able to:
1. Develop and apply critical thinking skills.
2. Design and present Lab as well as project reports
3. Apply appropriate methods for the analysis of raw data
4. Perform logical troubleshooting as and when required.
5. Verify and implement the concepts and theory learnt in class.

List of Experiments

1. Introduction to MySQL, Postgre Sql, Microsoft Sql softwares.
2. An exercise of data types in PostGresql & Data Definition Language Commands
3. Exercise on Data Manipulation Language and Transaction Control Commands using PostgreSql.
4. Exercise on Types of Data Constraints using PostgreSql.
5. Exercise on JOINS (Single-Table) Using Normalization
6. Exercise on JOINS (Multiple-Table) Using Normalization
7. Exercise on GROUP BY/ORDER BY Clause and Date Arithmetic using PostgreSql.
8. Exercise on different Functions (Aggregate, Math and String)
9. Exercise on different types of sub queries, Procedures, View and Triggers

ENERGY AND ENVIRONMENT			
Course Code:	**ES 415**	**Course Credits:**	**(L–T–P): 3 (3-0-0)**
Course Category:	**DSE**	**Course (C/P):**	**P**
Course Year (U/P):	**1P**	**Course Year (U/P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03+00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L+T)	**45+00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVE	
1.	To provide in-depth knowledge of renewable and non-renewable energy resources
2.	To provide knowledge about energy harnessing techniques
3.	To study the energy-environment issues
COURSE OUTCOME	
At the end of the course the students can use the knowledge so gathered and utilized to	
1.	Meet the challenges of energy *vis-a-vis* environmental security

UNIT I SUN AS SOURCE OF ENERGY

Nature of its radiation, solar radiation and its spectral characteristics; Conventional energy sources (coal, oil, biomass and natural gas), Non-conventional energy sources (hydro-electric power, tidal, wind, geothermal, solar, nuclear magneto-hydrodynamic power MHD); Energy use pattern in India and parts of world, Energy security.

UNIT II FOSSIL FUELS

Classification, composition, physico-chemical characteristics; Calorific value – gross and net; Energy content of coal, petroleum and natural gas, shale oil, coal bed methane, gas hydrates

Concept of Green Energy; Principles of generation of hydro-power, tidal energy, ocean thermal energy conversion, wind power, geothermal energy, solar energy (solar collectors, photo-voltaic modules, solar ponds).

UNIT III NUCLEAR ENERGY

Fission and fusion, Nuclear fuels, Nuclear reactor – principles and types; Mechanism of radiation action on living organisms - Stochastic and Non-stochastic effects, delayed effects; Radioactivity from nuclear reactors, fuel processing and radioactive waste, hazards related to power plants

UNIT IV BIOENERGY

Types, importance, methods of energy production from biomass

UNIT V ENVIRONMENTAL IMPLICATIONS OF ENERGY USE

CO2 emission and atmosphere –scenario in developed and developing world (and India), Global warming, Radiative forcing, Impacts of large scale exploitation of solar, wind, hydro, nuclear and bio-energy sources; National Solar Mission, National Mission for Enhanced Energy Efficiency, case studies.

SUGGESTED READINGS

Fay, J.A. and Golomb, D.S. 2011. *Energy and the Environment*, Oxford University Press, New Delhi. Iqbal, M. 1983. *An Introduction to Solar Radiation*. Academic Press, New York.

Kaushika, N.D. and Kaushik, K. 2004. *Energy, Ecology and Environment: A Technological Approach*, Capital Publications, New Delhi.

Website **-** *https://nptel.ac.in/course.html*

Semester-II

PYTHON PROGRAMMING			
Course Code:	**CD522**	**Course Credits:**	**3**
Course Category:	**CC**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester**	**2P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. To introduce students to the Python programming language.
2. To handle object oriented problems with Python code.
3. To Produce Python code to statistically analyse a dataset.
4. How to effectively visualize results.
5. An applied understanding of how to manipulate and analyze uncurated datasets

COURSE OUTCOMES
At the end of the course the students should be able to:
1. Define python environment and constructs of Python language.
2. Construct scripts in Python language.
3. Analyze data with Python Libraries.
4. Understanding the data, performing preprocessing, processing and data visualization to get insights from data.
5. Develop the model for data analysis and evaluate the model performance.

UNIT I

Python Introduction: Installing and setting Python environment in Windows and Linux, basics of Python interpreter, Execution of python program, Editor for Python code, syntax, variable, types. Flow control: if, if-else, for, while, range () function, continue, pass, break. Strings: Sequence operations, String Methods, Pattern Matching.

UNIT II

Lists: Basic Operations, Iteration, Indexing, Slicing and Matrixes; Dictionaries: Basic dictionary operations; Tuples: Basic Operations, Iteration, Indexing, Slicing; Functions: Definition, Call, Arguments, Scope rules and Name resolution;

UNIT III

Modules: Module Coding Basics, Importing Programs as Modules, Executing Modules as Scripts, Compiled Python files(.pyc), Standard Modules: OS and SYS, The dir() Function, Packages.

UNIT IV

Object Oriented Programming in Python: Classes, Objects, Inheritance, Operator Overloading, File Handling: Errors and Exceptions Handling (try and except) User-Defined Exception Objects, Regular expressions, User Defined Package with Python.

UNIT V

Python Packages for Data Sciences: Mathematical and Statistical Analysis with NumPy,and Pandas,

Manuplating and Visulisation of Data with SciPy, shaping, merging, reshaping, slicing datasets and Data structure with Pandas Library, 2d Plot with matplotlib and seaborn, Learning Package: sklearn

Text Books

1. Mark Lutz., Learning Python, Latest Edition, O'REILLY Media, Inc.
2. David Ascher and Mark Lutz, Learning Python, Publisher O'Reilly Media.
3. Reema Thareja, "Python Programming using Problem Solving approach",Oxford University Press.
4. Wes Mckinney "Python for Data Analysis", First edition, Publisher O'Reilly Media.

Reference Books

1. Allen Downey ,Jeffrey Elkner ,Chris Meyers,: Learning with Python, Dreamtech Press
2. David Taieb ,"Data Analysis with Python: A Modern Approach " 1st Edition, Packt Publishing.

AI ENABLED DATA SCIENCE			
Course Code:	**CD524**	**Course Credits:**	**3**
Course Category:	**CC7**	**Course (U / P)**	**P**
Course Year (U / P):	**2P**	**Course Semester**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. Become familiar with basic principles of AI toward problem solving, inference, perception, knowledge representation, and learning
2. Experience AI development tools such as an 'AI language', expert system shell, and/or data mining tool.
3. To extract valuable information for use in strategic decision making, product development, trend analysis, and forecasting.
4. To enable the students to understand about collection, presentation and analysis of data.
5 To study optimization algorithms with single and multi-variables for large datasets.

COURSE OUTCOMES
At the end of the course the students should be able to:
1. Demonstrate fundamental understanding of the history of artificial intelligence (AI) and its foundations.
2. Demonstrate proficiency developing applications in an 'AI language', expert system shell, or data mining tool.
3. Find the Eigen values and Eigenvectors of a matrix; enable to find maximum & minimum values of a function.
4. Having computational thinking (Ability to translate vast data into abstract concepts and to understand database reasoning.
5. Apply research- based knowledge to analyze and manage projects in data science.

UNIT I

Introduction to Artificial Intelligence and Search: Foundations and History of Artificial Intelligence, Applications of Artificial Intelligence, Intelligent Agents, Structure of Intelligent Agents. Computer vision, Natural Language Possessing.

Searching for solutions, Uniformed search strategies, Informed search strategies, Local search algorithms and optimistic problems, Adversarial Search, Search for games, Alpha - Beta pruning

UNIT II

Introduction to Data Science: Understanding Exploratory Data Analysis**,** Machine Learning**,** Model selection and evaluation**,** Data Warehousing**,** Data Mining**,** Data Visualization**,** Cloud Computing**,** Business Intelligence**,** Storytelling with Data**,** Communication and Presentation.

UNIT III

Statistics for Data Science: Sampling Techniques, Data Classification, Tabulation, Frequency and graphic Representation, Measures of Central Tendency, Measures of Variation, Quartiles and Percentiles – Moments, Skewness and Kurtosis. Scatter Diagram, Karl Pearson's Correlation Coefficient, Rank Correlation, Correlation Coefficient for Bi-variate Frequency Distribution, Regression Coefficients, Fitting of Regression Lines.

UNIT IV

Stochastic Models and Numerical Optimization: Random variables and events, distributions, inequalities and limits, Stochastic processes, Exponential distribution, Markov chains, Discrete Time Markov chains, Continuous Time Markov chains, Mathematical models of optimization.

UNIT V

Big Data Analytics Through Hadoop: Big Data and its importance, Sources of Big Data, Characteristics of Big Data, Big Data Analytics, Big Data Applications. Hadoop Distributed File System, Map Reduce Paradigm, Moving Data in and out of Hadoop, Understanding inputs and outputs of Map Reduce, Data Serialization. Hadoop architecture, Hadoop ecosystem. HDFS Federation, Map Reduce, yarn.

Suggested Text Books

1. Jojo Moolayil, "Smarter Decisions : The Intersection of IoT and Data Science", PACKT, 2016.
2. Cathy O'Neil and Rachel Schutt , "Doing Data Science", O'Reilly, 2015.
3. David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big data Analytics", EMC 2013
4. Raj, Pethuru, "Handbook of Research on Cloud Infrastructures for Big Data Analytics", IGI Global.
5. Gupta, S.C. and Kapoor, V.K.: "Fundamentals of Mathematical Statistics", Sultan & Chand & Sons, New Delhi, 11th Ed, 2002.
6. Hastie, Trevor, et al. "The elements of Statistical Learning", Springer, 2009

OPTIMIZATION TECHNIQUES FOR DATA SCIENCE			
Course Code:	**CD526**	**Course Credits:**	**3**
Course Category:	**CC**	**Course (U / P)**	**P**
Course Year (U / P):	**2P**	**Course Semester**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. To learn how to analyze and solve a linear system of equations in context of data science.
2. To understand important characteristics of matrices, such as its four fundamental subspaces, rank, determinant, Eigen values and Eigen vectors.
3. To learn concepts of vector spaces such as independence, basis, dimensions, orthogonality.
4. To enable the students to understand about collection, presentation and analysis of data.
5 To study optimization algorithms with single and multi-variables for large datasets.

COURSE OUTCOMES
At the end of the course the students should be able to:
1. Understand the basic concepts of inner product space, norm, angle, Orthogonality and projection and implementing the Gram-Schmidt process, to obtain Least square solution and SVD.
2. Understand the basic concepts of vector space, subspace, basis and dimension.
3. Find the Eigen values and Eigenvectors of a matrix, enable to find maximum & minimum values of a function.
4. Use classical optimization techniques and numerical methods of optimization.
5. Apply research- based knowledge to analyze and manage projects in data science.

UNIT I

Linear Algebra and Optimization: Vector spaces, subspaces, bases and dimensions, systems of linear equations, Linear transformations, Isomorphism, Inner product, Orthogonality, Eigen values and Eigenvectors, Matrix factorizations, Function optimization, Newton's method.

UNIT II

Introduction to search: Conditions for local minimization - One dimensional Search methods - Golden search method, Fibonacci method, Newton's Method, Secant Method, Remarks on Line Search Gradient-based methods - introduction, the method of steepest descent, analysis of Gradient Methods, Convergence, Convergence Rate, Analysis of Newton's Method, Newton's Method for Nonlinear Least-Squares - Conjugate direction method, Conjugate Direction Algorithm, Conjugate Gradient Algorithm for Non-Quadratic Quasi Newton method.

UNIT III

Principles of Artificial Intelligence and Machine Learning: Intelligent Agents, Problem solving and search, Uninformed search, Knowledge and reasoning, Probabilistic reasoning, Bayesian networks and decision theory, Neural networks, Issues in ANN training, Types of ANN architectures, SVM.

UNIT IV

Stochastic Models and Numerical Optimization: Random variables and events, distributions, inequalities and limits, Stochastic processes, Exponential distribution, Markov chains, Discrete Time Markov chains, Continuous Time Markov chains, Mathematical models of optimization.

UNIT V

Exploration and Statistical Analysis for Data Science: Data Science process, Memorization methods, unsupervised models, Univariate data exploration, Data visualization, Prediction and filtering, Probability theory and Statistics.

Suggested Text Books

1. Howard Anton, Chris Rorres, Elementary Linear Algebra, Tenth edition, John Wiley & Sons, 2010
2. 2.Edwin K.P. Chong, Stanislaw H. Zak, An introduction to Optimization, Second edition, Wiley, 2013
3. Nabil Nassif, Jocelyne Erhel, Bernard Philippe, Introduction to Computational Linear Algebra, CRC press, 2015
4. Gilbert Strang, Linear Algebra and Its Applications, Fourth edition, Cengage, 2006
5. Mohan C. Joshi andKannan M. Moudgalya, Optimization: Theory and Practice, Narosa Pub- lishing House, New Delhi, 2004
6. Hal Daum III, A Course in Machine Learning, 2015

ADVANCED MS EXCEL			
Course Code:	**CD528**	**Course Credits:**	**3**
Course Category:	**CC**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester**	**2P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. Understand the basics and functions of MS excel.
2. Clear understanding and use of data validations and templates.
3. Purpose of sorting and filtering features.
4. Use of reports in business organizations.
5. Purpose and advantage of charts for top management in any work place.
COURSE OUTCOMES
At the end of the course the students should be able to:
1. Learn to understand the functions in Excel.
2. Understand the validations.
3. Make reports in excel.
4. Learn to work with pivot tables.
5. Learn how to make charts in MS excel.

UNIT I Overview of the Basics of Excel

Customizing common options in Excel, Absolute and relative cells, Protecting and un-protecting worksheets and cells, Writing conditional expressions (using IF), Using logical functions (AND, OR, NOT), Using lookup and reference functions (VLOOKUP, HLOOKUP, MATCH, INDEX), VlookUPwith Exact Match, Approximate Match, Nested VlookUP with Exact Match, VlookUP with Tables, Dynamic Ranges, Nested VlookUP with Exact Match, Using VLookUP to consolidate Data from Multiple Sheets.

UNIT II Data Validations& Templates

Specifying a valid range of values for a cell, specifying a list of valid values for a cell, specifying custom validations based on formula for a cell, designing the structure of a template, Using templates for standardization of worksheets.

UNIT III Sorting and Filtering Data and Working with Reports

Sorting tables, Using multiple-level sorting, Using custom sorting, Filtering data for selected view (AutoFilter), Using advanced filter options, Creating subtotals, Multiple-level subtotals

UNIT IV Pivot tables

Formatting and customizing Pivot tables, Using advanced options of Pivot tables, Pivot charts Consolidating data from multiple sheets and files using Pivot tables, Using external data sources, Using data consolidation feature to consolidate data, Show Value As (% of Row, % of Column, Running Total, Compare with Specific Field), Viewing Subtotal under Pivot, Creating Slicers

UNIT V Charts

Using Charts, Formatting Charts, Using 3D Graphs, Using Bar and Line Chart together, Using Secondary Axis in Graphs, Sharing Charts with PowerPoint / MS Word, Dynamically, (Data Modified in Excel, Chart would automatically get updated)

Text Books

1. Excel 2016 Bible by Walkenbach John, John WIley & Sons publications.
2. Excel: Quick Start Guide from Beginner to Expert, by William Fischer

Reference Books

3. Power Pivot and Power BI, by Rob Collie and Avichal Singh

PARALLEL AND DISTRIBUTED COMPUTING			
Course Code:	**CD534**	**Course Credits:**	**3**
Course Category:	**E1**	**Course (U / P)**	**P**
Course Year (U / P):	**2P**	**Course Semester**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. To learn parallel and distributed algorithms development techniques for shared memory and message passing models.
2. To study the main classes of parallel algorithms.
3. To study the complexity and correctness models for parallel algorithms.
4. To introduce fundamental principles of distributed systems, technical challenges and key design issues.
5. To impart knowledge of the distributed computing models, algorithms and the design of distributed system.
COURSE OUTCOMES
At the end of the course the students should be able to:
1. Understand, appreciate and apply parallel and distributed algorithms in problem Solving.
2. Evaluate the impact of network topology on parallel/distributed algorithm formulations and traffic their performance.
3. Master skills to measure the performance of parallel and distributed programs.
4. To learn and apply knowledge of parallel and distributed computing techniques and methodologies.
5. To gain experience in the design, development, and performance analysis of parallel and distributed applications.

UNIT I

Introduction: Parallel Processing, Shared Memory Multiprocessing, Distributed Shared Memory, Message Passing Parallel Computers.

Processes & Shared Memory Programming Processes: Shared Memory Programming, General Model Of Shared Memory Programming, Forking-Creating Processes, Joining Processes, Process Model Under UNIX.

UNIT II

Basic Parallel Programming Techniques: Loop Splitting, Ideal Speedup, Spin-Locks, Contention and Self- Scheduling.

Scheduling: Loop Scheduling, Variations On Loop Scheduling, Self- Scheduling, Variations On Self-Scheduling, Indirect Scheduling, Block Scheduling. Thread based Implementation, Programming Using the Message Passing Paradigm.

UNIT III

Evolution of Distributed Computing - Issues in designing a distributed system, Challenges, Minicomputer

model, Workstation model, Workstation-Server model, Processor, pool model, Trends in distributed systems.
System models: Physical models, Architectural models, Fundamental models.

UNIT IV

Interprocess communication: characteristics, group communication, Multicast Communication Remote Procedure call , Network virtualization. Case study: Skype.
Distributed file system: File service architecture, Network file system, Andrew file system- Name Service.

UNIT V

Synchronous Computations, load balancing, distributed termination examples, programming with shared memory, shared memory multiprocessor constructs for specifying parallelist sharing data parallel programming languages and constructs, open MP.
Distributed shared memory systems and programming achieving constant memory distributed shared memory programming primitives, algorithms – sorting and numerical algorithms.

Suggested Text Books

1. George Coulouris, Jean Dollimore and Tim Kindberg , Distributed Systems: Concepts and Design, Fifth Edition , Pearson Education, 2011
2. Pradeep K Sinha, Distributed Operating Systems : Concepts and Design, Prentice Hall of India
3. Sukumar Ghosh, Distributed Systems: An Algorithmic Approach, Second Edition
4. A.D. Kshemkalyani, M. Singhal, Distributed Computing: Principles, Algorithms, and Systems, ISBN: 9780521189842, Cambridge University Press, March 2011.
5. Introduction to Parallel algorithms by Jaja from Pearson, 1992.

PREDICTIVE ANALYTICS AND DATA VISUALISATION			
Course Code:	**CD538**	**Course Credits:**	**3**
Course Category:	**E2**	**Course (U / P)**	**P**
Course Year (U / P):	**2P**	**Course Semester**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. To learn, how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines and Bayesian network models.
2. To advice on when and how to use each model. Also learn how to combine two or more models to improve prediction
3. To extend student's knowledge in the area of Data Science with emphasis on Predictions utilizing associated statistical methods and software tools.
4. To introduce fundamental principles of distributed systems, technical challenges and key design issues.
5. Choose appropriate types and formats of data for topical, network, burst, and temporal analysis. Navigate to data sources. Download data in proper format.
COURSE OUTCOMES
At the end of the course the students should be able to:
1. Understand the process of formulating business objectives, data selection/collection, preparation and process to successfully design, build, evaluate and implement predictive models for a various business application.
2. Select appropriate predictive modeling approaches to identify cases to progress with.
3. Ability to apply specific statistical and regression analysis methods applicable to predictive analytics to identify new trends and patterns, uncover relationships, create
4. Ability to develop and use various quantitative and classification predictive models based on various regression and decision tree methods.
5. To gain experience in the design, development, and performance analysis of parallel and distributed applications.

UNIT I

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

Model Assesment and Selection : Bias,Variance,and model complexity,Bias-variance trade off, Optimisim of the training error rate ,Esimate of In-sample prediction error,Effective number of parameters, Bayesian approach and BIC, Cross- validation ,Boot strap methods, conditional or

UNIT II

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

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

UNIT III

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

UNIT IV

Visualization: Introduction of visual perception, visual representation of data, Gestalt principles, information overloads. Creating visual representations, visualization reference model, visual mapping, and visual analytics. Design of the visualization applications.

Data Visualization Tools: Rank Analysis Tools- Trend Analysis Tools- Multivariate Analysis Tools- Distribution Analysis Tools- Correlation Analysis Tools- Geographical Analysis Tools.

UNIT V

Classification of visualization & Types: Classification of visualization systems, Interaction and visualization, techniques misleading, Visualization of one, two and multi-dimensional data, text and text documents. Classification of visualization systems, Interaction and visualization techniques misleaing.

Suggested Text Books

1. Andy Kirk, Data Visualization A Handbook for Data Driven Design, Sage Publications, 2016
2. Philipp K. Janert, Gnuplot in Action, Understanding Data with Graphs, Manning Publications, 2010.
3. Alberto Cordoba, “Understanding the Predictive Analytics Lifecycle”, Wiley, 2014.
4. Eric Siegel, Thomas H. Davenport, “Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die”, Wiley, 2013.
5. James R Evans, “Business Analytics – Methods, Models and Decisions”, Pearson 2013.
6. R. N. Prasad, S eema Acharya, “Fundamentals of Business Analytics”, Wiley, 2015.

PYTHON PROGRAMMING LAB			
Course Code:	**CD582**	**Course Credits:**	**2**
Course Category:	**CC-L3**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester**	**2P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1.To understand the basic constructs of Python Interpreter.
2.To demonstrate the working of Python functions and modules w.r.t definition call and scope.
3. To make acquainted with OOPS and File handling concept in Python.
4.To understand and apply various Python packages for Data handling.
5. To understand how to visualize the data.
COURSE OUTCOMES
At the end of the course the students should be able to:
1.Write and demonstrate script in Python by using basic constructs and control statements of Python.
2. Illustrate the use of OOPS and file handling concept for data handling and visualization.
3. Synthesize the code in Python by making a use of various Data Handling libraries.
4. Illustrate the working of data visualization.

List of Programs:

1. Write a Python program to reverse the internal content of the words for given sentence and also count the reversed words, the console output should be in comma-separated manner.

2. Write a Python program to find the list of words that are longer than n from a given list of words.

3. Write a Python program to combine two dictionary adding values for common keys. d1 = {'a': 100, 'b': 200, 'c':300}
d2 = {'a': 300, 'b': 200, 'd':400}
Sample output: Counter({'a': 400, 'b': 400, 'd': 400, 'c': 300})

4. Write a python program to demonstrate a dictionary in tabular format as shown in given below.

Col_1	Col_2	Col_3
101	501	989

204	698	102
329	754	141

5. Write a Python program to write a Python dictionary to a csv file. After writing the CSV file read the CSV file and display the content.

6. Write a Python program to merge two different datasets based on the common attributes and determine NULL values (if any).

7. Write a Python program to display the result Year wise (like 2018, 2019) from the given below dataset and also display the result using graphs.

Full Name	Age	Height	Address	position	Timestamp
Hacker	20	176.5	Hyderabad	Developer	02/02/202019:20
Smith	24	171.5	Ranchi	SystemAnalyst	05/09/201911:52
Beatriz	21	163.5	Kerala	BlogWriter	03/01/202022:27
Devin	28	156	UttarPradesh	Developer	17/04/201920:33:07
Aaron	23	156	Delhi	BlogWriter	02/10/201910:36
Diego	29	171	Patna	DataScientist	27/01/202023:15:46
Driss	23	160.5	Kanpur	SoftwareDeveloper	07/06/201916:35
Shira	26	162	Chennai	PythonProgrammer	13/10/202019:19:37
Fumio	23	156	Yaswantpur	JavaDeveloper	12/02/202010:34
Amara	25	163	Nagpur	MachineLearningE ngg.	19/03/202020:54:32
Chiyo	23	145	Pune	BlogWriter	04/12/201919:13
Zahra	27	156.5	Noida	PhpDeveloper	03/10/201920:24
Faina	24	167	Gurugram	DataScientist	27/10/202017:40:57
Penelope	23	156	Haryana	DataEngineer	24/10/202006:44:06
Ella	18	159	J&K	Teacher	29/04/201920:14:20
Alice	26	159	Noida	C++Developer	24/07/202014:54:36
John	23	164.5	Bangluru	C/C++Developer	22/09/201901:28:58
Bob	22	170.5	Mumbai	Bussinessman	08/06/201911:22
Jackel	30	158.5	Karnataka	Adminstrative	07/03/20196:56
Evie	23	156	Guntur	HR	19/03/202011:51:41

8. Write a Python program to create an instance of a specified class and display the namespace of the said instance.

9. Write a Python class named Student with two attributes student_id, student_name. Add a new attribute student_class. Create a function to display the entire attribute and their values in Student class.

10. Write a Python program to read a csv file and perform various operations using pandas and numpy packages.

ADVANCED MS EXCEL LAB			
Course Code:	**CD584**	**Course Credits:**	**2**
Course Category:	**CC-L4**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**2P**
No. of Lectures + Tutorials +Lab (Hrs/Week):	**0+ 0+3**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T+P):	**0+ 0+ 10**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. Understand the basics and functions of MS excel.
2. Clear understanding and use of data validations and templates.
3. Purpose of sorting and filtering features.
4. Use of reports in business organizations.
5. Purpose and advantage of charts for top management in any work place.

COURSE OUTCOME
At the end of the course the students should be able to:
1. Learn to understand the functions in Excel.
2. Understand the validations.
3. Make reports in excel.
4. Learn to work with pivot tables.
5. Learn how to make charts in MS excel.

List of Experiments

1. Create a new workbook and save the file with the name "Payroll". Enter the labels and values in the exact cells locations as desired. Use AutoFill to put the Employee Numbers into cells A6:A8. Set the columns width and rows height appropriately.

2. Create a workbook and enter relatable data of some employees. Calculate the Gross Pay for employee; enter a formula in cell E4 to multiply Hourly Rate by Hours Worked. Calculate the Social Security Tax (S.S Tax), which is 6% of the Gross Pay; enter a formula in cell F4 to multiply Gross Pay by 6%.

3. Create a workbook. Enter data as required. Calculate the Net Pay; enter a formula in cell G4 to subtract Social Security Tax from Gross Pay. Set the work sheet vertically and horizontally on the page.

4. Create a workbook having relatable data of sales of various models of cars in a showroom. Create a 3-dimensional column chart comparing sales data for men and women sales person.

5. Create a pie chart to compare the favourite films data for 15-25 year olds only (be careful not to include any unnecessary blanks rows or columns in your selected data).

6. Create a pivot table from this data, then use the filters within to view the average prices of holidays that have a Travel Method of Plane and a Resort Name that begins with the letter S.

Wise Owl Travel Agents

Country	Resort Name	No of Days	Travel Method	Price	Holiday ID
Australia	Great Barrier Reef	32	Plane	£750	I990AUS
Australia	Perth	28	Plane	£985	AUS112J
Chile	Santiago	21	Plane	£1,259	CH266H
England	London	3	Train	£69	I456UK
England	Bognor	1	Coach	£12	BG726H
France	Lyon	14	Plane	£399	A7995FR
France	Paris - Euro Disney	5	Train	£269	TH789FR
France	Paris - Euro Disney	3	Train	£125	TH788FR

7. Create an If function to calculate whether each movie was a flop or a success. Use the following criteria: If the profit was less than 100,000,000 then the movie is a flop otherwise the movie is a success.

8. Create an If function to rate the players based on the following criteria: If a player scores more than 15 points he has a High score otherwise he must try harder.

9. Convert this data into a pivot table and find the overall average speed of all rides that satisfy the following criteria: The Type is Steel, The Design is Sit Down, The Amusement Park has the word adventure somewhere in the title

Roller Coaster	Amusement Park	Type	Design	Status	Opened	Speed (mph)
Air	Alton Towers	Steel	Flying	Operating	2002	46.6
Boomerang	Pleasure Island Family Theme Park	Steel	Sit Down	Operating	1993	47
Cobra	Paultons Park	Steel	Sit Down	Operating	2006	31.1
Colossus	Thorpe Park	Steel	Sit Down	Operating	2002	45
Corkscrew	Alton Towers	Steel	Sit Down	Operating	1980	40
Corkscrew	Flamingo Land Theme Park & Zoo	Steel	Sit Down	Operating	1983	40
Crazy Mouse	South Pier	Steel	Sit Down	Operating	1998	29.1
Crazy Mouse	Brighton Pier	Steel	Sit Down	Operating	2000	29.1
Enigma	Pleasurewood Hills	Steel	Sit Down	Operating	1995	34
Express	M&Ds Scotland's Theme Park	Steel	Sit Down	Operating	2006	28
Fantasy Mouse	Fantasy Island	Steel	Sit Down	Operating	2000	29.1

10. Create a worksheet and add desired data. Find TAX (If ITEM PRICE is less than 100, TAX is 50, otherwise it should be 100). TOTAL PRICE BEFORE TAX =NO. OF ITEMS * ITEM PRICE. TOTAL PRICE AFTER TAX = TOTAL PRICE BEFORE TAX + TAX. RATE (If TOTAL PRICE AFTER TAX > 3500 then the rate is “HIGH”, otherwise it is REASONABLE. Find Count of Items, Average of Taxes, Min Item PRICE and Max Item PRICE.

Semester III

MACHINE LEARNING ALGORITHMS			
Course Code:	**CD621**	**Course Credits:**	**3**
Course Category:	**CC**	**Course (U / P):**	**P**
Course Year (U / P):	**2P**	**Course Semester :**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. To explain the concept of how to learn patterns and concepts from data without being explicitly programmed in various IOT nodes.
2. To design and analyze various machine learning algorithms and techniques with a modern outlook focusing on recent advances.
3. Explore supervised and unsupervised learning paradigms of machine learning.
4. To explore Deep learning technique and various feature extraction strategies.
5. Acquire Data Analysis skills, and Create AI/ML solutions for various business problems.
COURSE OUTCOMES
At the end of the course the students should be able to:
1. Recognize the characteristics of machine learning that make it useful to real-world problems.
2. To compare and contrast pros and cons of various machine learning techniques and to get an insight of when to apply a particular machine learning approach.
3. To mathematically analyze various machine learning approaches and paradigms.
4. Characterize machine learning algorithms as supervised, semi-supervised, and unsupervised.
5. Understand the concept behind neural networks.

UNIT I

Introduction to learning Techniques: Supervised Learning (Regression/Classification), Basic methods: Distance-based methods, Nearest-Neighbours, Decision Trees, Naive Bayes, Linear models: Linear Regression, Logistic Regression, Generalized Linear Models, Support Vector Machines, Nonlinearity and Kernel Methods, Beyond Binary Classification: Multi-class/Structured Outputs, Ranking

UNIT II

Unsupervised Learning: Clustering: K-means/Kernel K-means, Dimensionality Reduction: PCA and kernel PCA, Matrix Factorization and Matrix Completion, Generative Models (mixture models and latent factor models)

UNIT III

Evaluating Machine Learning algorithms and Model Selection, Introduction to Statistical Learning Theory, Ensemble Methods (Boosting, Bagging, Random Forests).

Sparse Modeling and Estimation, Modeling Sequence/Time-Series Data, Deep Learning and Feature Representation Learning.

UNIT IV

Neurons and biological motivation, linear threshold units, Perceptrons: representational limitation and gradient descent training, Multilayer networks and backpropagation, Hidden layers and constructing intermediate, distributed representations, Overfitting, learning network structure, recurrent networks.

UNIT V

Scalable Machine Learning (Online and Distributed Learning) A selection from some other advanced topics, e.g., Semi-supervised Learning, Active Learning, Reinforcement Learning, Inference in Graphical Models, Introduction to Bayesian Learning and Inference.

Introduction to Simulation Tool for Machine Learning (like WEKA, R, MATLAB)

Recent trends in various learning techniques of machine learning and classification methods.

Text and References Books:

1. Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press.

2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Springer.

3. Christopher Bishop, Pattern Recognition and Machine Learning, Springer.

4. Shai Shalev-Shwartz, Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press.

BIG DATA FRAMEWORK FOR DATA SCIENCE			
Course Code:	**CD623**	**Course Credits:**	**3**
Course Category:	**CC**	**Course (U / P):**	**P**
Course Year (U / P):	**2P**	**Course Semester :**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. To prepare the Big Data for analysis.
2. To extract the meaningful data from unstructured Big Data.
3. To develop Data Visualizations skill.
4. To apply various tools for analysis of structured data.
5. To apply various tools for analysis of unstructured Big Data.
COURSE OUTCOMES
At the end of the course the students should be able to:
1. Able to analyze the identification of Big Data problem
2. Able to extract the structured data from unstructured data.
3. Use Hadoop related tools such as JAQL.
4. Use Spark, Pig and Hive for structured Data analytics
5. Use Spark, Pig and Hive for unstructured Big Data analytics.

UNIT I Big Data Introduction

What is big data, why big data, convergence of key trends, unstructured data, industry examples of big data, web analytics, big data and marketing, fraud and big data, risk and big data, big data and healthcare, big data in medicine, advertising and big data, big data technologies, open source technologies, cloud and big data, mobile business intelligence, Crowd sourcing analytics, inter and trans firewall analytics. Data Gathering and Preparation: Data formats, parsing and transformation, Scalability and real-time issues.

UNIT II Data Cleaning

Consistency checking, Heterogeneous and missing data, Data Transformation and segmentation. Visualization: Descriptive and comparative statistics, Designing visualizations, Time series, Geo-located data, Correlations and connections, Hierarchies and networks, interactivity.

UNIT III Big Data Technology

Big Data Architecture, Big Data Warehouse, Functional Vs. Procedural Programming Models for Big Data NoSQL: Introduction to NoSQL, aggregate data models, key-value and document data models.

UNIT IV Big Data Tools

Introduction to Hadoop Ecosystem, Hadoop, Requirement of Hadoop Framework, Design principle of Hadoop, Comparison with other system, Hadoop Components, Hadoop 1 vs Hadoop 2, Hadoop Daemon's, HDFS Commands, Map Reduce Programming: I/O formats, Map side join, Reduce Side Join, Secondary sorting, Pipelining Map Reduce jobs

UNIT V Advance Big Data Tools
Spark, PIG, JAQL, Understanding Text Analytics and Big Data, Predictive Analysis of Big Data, Role of Data Analyst.

Text and References Books:
1. EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, John WIley & Sons.
2. Anil Maheshwari, Data Analytics Make Accesible, Orilley Publications.
3. Croll and B. Yoskovitz Lean Analytics: Use Data to Build a Better Startup Faster, Oreilley Publications.

Big Data Framework for Data Science Lab			
Course Code:	**CD681**	**Course Credits:**	**2**
Course Category:	**CC-L**	**Course (U / P)**	**P**
Course Year (U / P):	**2P**	**Course Semester (U / P):**	**3P**
No. of Lectures + Tutorials +Lab (Hrs/Week):	**3**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T+P):	**10**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. Understand and implement the basics of data structures in Java.
2. Demonstrate the knowledge of big data analytics and implement different file management task in Hadoop
3. Understand Big Data for Business Intelligence.
4. Learn business case studies for Big Data Analytics.
5. Analyze and perform different operations on Hadoop.
COURSE OUTCOME
At the end of the course the students should be able to:
1. Understand and implement the basics of data structures in Java.
2. Demonstrate the knowledge of Big Data Analytics and implement different file management task in Hadoop.
3. Understand Map Reduce Paradigm and develop data applications using variety of Systems.
4. Analyze and perform different operations on data using Pig Latin Scripts.
5. Illustrate and apply different operations on relations and databases using Hive.

List of Experiments:

1. Implement the following Data structures in Java
i) Linked Lists, ii) Stacks, iii) Queues, iv) Set, v) Map

2. Perform setting up and Installing Hadoop in its three operating modes:
a) Standalone, b) Pseudo distributed, c) Fully distributed.

3. Implement the following file management tasks in Hadoop:
Adding files and directories, Retrieving files

4. Implement the following file management tasks in Hadoop:
Deleting files Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities.

5. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.

6. Write a Map Reduce program that mines weather data. Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented.

7. Implement Matrix Multiplication with Hadoop Map Reduce.

8. Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data.

9. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.

10. Solve some real life big data.

BLOCK CHAIN TECHNOLOGY FOR DATA SCIENCE			
Course Code:	**CS631**	**Course Credits:**	**3**
Course Category:	**E**	**Course (U / P)**	**P**
Course Year (U / P):	**2P**	**Course Semester**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**
COURSE OBJECTIVES			
1. To understand the technology behind blockchain			
2. Explain distributed Consensus, and Consensus in Bitcoin			
3. Discuss Permissioned Blockchain, and Hyperledger Fabric			
4. To comprehend the issues related to blockchain			
5. To study the real-world applications of blockchain			
COURSE OUTCOMES			
After completion of course, students would be able to:			
1. Describe the basic concept of Blockchain, Crypto Primitives, Bitcoin Basics			
2. Identify the area in which they can apply permission or permission less blockchain.			
3. Apply Block chaining concept in various applications.			
4. Design and implement new ways of using blockchain for applications other than cryptocurrency			
5. Recognize the underlying technology of transactions, blocks, proof-of-work, and consensus building			

UNIT I

Introduction to Blockchain: What is Blockchain, Public Ledgers, Blockchain as Public Ledgers, Bitcoin, Blockchain 2.0, Smart Contracts, Block in a Blockchain, Transactions, Distributed Consensus, The Chain and the Longest Chain, Cryptocurrency to Blockchain 2.0, Permissioned Model of Blockchain

UNIT II 15

Basic Crypto Primitives: Cryptographic Hash Function, Properties of a hash function, Hash pointer and Merkle tree, Digital Signature, Public Key Cryptography, A basic cryptocurrency.

Bitcoin Basics: Creation of coins, Payments and double spending, FORTH – the precursor for Bitcoin scripting, Bitcoin Scripts, Bitcoin P2P Network, Transaction in Bitcoin Network, Block Mining, Block propagation and block relay.

UNIT III

Distributed Consensus: Why Consensus, Distributed consensus in open environments, Consensus in a Bitcoin network.

Consensus in Bitcoin: Bitcoin Consensus, Proof of Work (PoW) – basic introduction, Hashcash PoW, Bitcoin PoW, Attacks on PoW and the monopoly problem, Proof of Stake, Proof of Burn and Proof of Elapsed Time. The life of a Bitcoin Miner, Mining Difficulty, Mining Pool.

UNIT IV

Permissioned Blockchain: Permissioned model and use cases, Design issues for Permissioned blockchains, Execute contracts, State machine replication, Consensus models for permissioned blockchain, Distributed consensus in closed environment, Paxos, RAFT Consensus, Byzantine general problem.

Blockchain Components and Concepts: Actors in a Blockchain, Components in Blockchain design, Ledger in Blockchain.

UNIT V

Hyperledger Fabric – Transaction Flow: Fabric Architecture, Transaction flow in Fabric.

Hyperledger Fabric Details: Ordering Services, Channels in Fabric, Fabric Peer and Certificate Authority.

Fabric – Membership and Identity Management: Organization and Consortium Network, Membership Service Provide, Transaction Signing.

Suggested Readings:

1. Nitin Gaur, Luc Desrosiers, Venkatraman Ramakrishna, Petr Novotny, Salman Baset, Anthony O'Dowd.Hands-On Blockchain with Hyperledger: Building decentralized applications with Hyperledger Fabric and Composer. Packt Publishing Ltd.

2. Bellaj Badr, Richard Horrocks, Xun (Brian) Wu. Blockchain By Example: A developer's guide to creating decentralized applications using Bitcoin, Ethereum, and Hyperledger. Packt Publishing Ltd, 2018.

3. Vikram Dhillon, David Metcalf, Max Hooper. Blockchain Enabled Applications: Understand the Blockchain Ecosystem and How to Make it Work for You. Apress.

4. Mayukh Mukhopadhyay Ethereum Smart Contract Development: Build blockchain-based decentralized applications using solidity. Packt Publishing Ltd.

ELECTIVES FROM DCSE

Data Security			
Course Code:	**CD532**	**Course Credits:**	**3**
Course Category:	**CC**	**Course (U / P)**	**U**
Course Year (U / P):	**4U**	**Course Semester (U / P):**	**7U**
No. of Lectures + Tutorials (Hrs/Week):	**03 +0+0**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45**	**End Sem. Exam Hours:**	**3**
COURSE OBJECTIVES			
1. To study the different models involved in data security			
2. To understand the security issues and solution for database			
3. To study application in real time world to protect the database and information			
4. Solve complex problems in a team of database work			
5. Identify security threats in database systems			
COURSE OUTCOMES			
At the end of the course the students should be able to:			
1. Avoid unauthorized data observation			
2. Ensure the data confidentiality			
3. Prove that data integrity is preserved			
4. Design and Implement secure database systems			
5. Avoid unauthorized data modification			

UNIT I INTRODUCTION TO DATABASE

Introduction to Database – Relational Database & Management System – ACID Properties, Normalization, RAID, Relational Algebra, Query tree, Data Abstraction (Physical Level, Logical Level & View Level) - Multi-level Database, Distributed Database

UNIT II SECURITY ISSUES

Security issues in Database – Polyinstantiation - Integrity Lock - Sensitivity Lock – Security Models – Access Control (Grant & Revoke Privileges) - Statistical Database, Differential Privacy. Distributed Database Security.

UNIT III OUTSOURCED DATABASE AND SECURITY REQUIREMENTS

Outsourced Database and security requirements – Query Authentication Dimension – Condensed RSA, Merkle Tree, B+ Tree with Integrity and Embedded Merkle B-Tree – Partitioning & Mapping - Keyword Search on Encrypted Data

UNIT IV PRIVACY PRESERVING DATA MINING

Privacy-Preserving Data Mining – Introduction - Randomization method: Privacy Quantification, Attacks on Randomization, Multiplicative Perturbations, Data Swapping - KAnonymity framework – Distributed Privacy-Preserving Data Mining.

UNIT V DATABASE WATERMARKING

Database Watermarking – Basic Watermarking Process - Discrete Data, Multimedia, and Relational Data – Attacks on Watermarking - Single Bit Watermarking, Multi bit Waterm

Reference Books

1. Michael Gertz and Sushil Jajodia (Editors), Handbook of Database Security: Applications and Trends , ISBN-10: 0387485325. Springer, 2007

2. Osama S. Faragallah, El-Sayed M. El-Rabaie, Fathi E. Abd El-Samie, Ahmed I. Sallam, and Hala S. El-Sayed, Multilevel Security for Relational Databases by; ISBN 978-1-4822- 0539-8. CRC Press, 2014.

3. Bhavani Thuraisingham, Database and Applications Security: Integrating Information Security and Data Management, CRC Press, Taylor & Francis Group, 2000

PARALLEL AND DISTRIBUTED COMPUTING			
Course Code:	**CD534**	**Course Credits:**	**3**
Course Category:	**E1**	**Course (U / P)**	**P**
Course Year (U / P):	**2P**	**Course Semester**	**3P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES

1. To learn parallel and distributed algorithms development techniques for shared memory and message passing models.
2. To study the main classes of parallel algorithms.
3. To study the complexity and correctness models for parallel algorithms.
4. To introduce fundamental principles of distributed systems, technical challenges and key design issues.
5. To impart knowledge of the distributed computing models, algorithms and the design of distributed system.

COURSE OUTCOMES

At the end of the course the students should be able to:
1. Understand, appreciate and apply parallel and distributed algorithms in problem Solving.
2. Evaluate the impact of network topology on parallel/distributed algorithm formulations and traffic their performance.
3. Master skills to measure the performance of parallel and distributed programs.
4. To learn and apply knowledge of parallel and distributed computing techniques and methodologies.
5. To gain experience in the design, development, and performance analysis of parallel and distributed applications.

UNIT I

Introduction: Parallel Processing, Shared Memory Multiprocessing, Distributed Shared Memory, Message Passing Parallel Computers.

Processes & Shared Memory Programming Processes: Shared Memory Programming, General Model Of Shared Memory Programming, Forking-Creating Processes, Joining Processes, Process Model Under UNIX.

UNIT II

Basic Parallel Programming Techniques: Loop Splitting, Ideal Speedup, Spin-Locks, Contention and Self- Scheduling.

Scheduling: Loop Scheduling, Variations On Loop Scheduling, Self- Scheduling, Variations On Self-Scheduling, Indirect Scheduling, Block Scheduling. Thread based Implementation, Programming Using the Message Passing Paradigm.

UNIT III

Evolution of Distributed Computing - Issues in designing a distributed system, Challenges, Minicomputer model, Workstation model, Workstation-Server model, Processor, pool model, Trends in distributed systems.

System models: Physical models, Architectural models, Fundamental models.

UNIT IV

Interprocess communication: characteristics, group communication, Multicast Communication Remote Procedure call , Network virtualization. Case study: Skype.

Distributed file system: File service architecture, Network file system, Andrew file system- Name

Service.

UNIT V

Synchronous Computations, load balancing, distributed termination examples, programming with shared memory, shared memory multiprocessor constructs for specifying parallelist sharing data parallel programming languages and constructs, open MP.

Distributed shared memory systems and programming achieving constant memory distributed shared memory programming primitives, algorithms – sorting and numerical algorithms.

Suggested Text Books

6. George Coulouris, Jean Dollimore and Tim Kindberg , Distributed Systems: Concepts and Design, Fifth Edition , Pearson Education, 2011
7. Pradeep K Sinha, Distributed Operating Systems : Concepts and Design, Prentice Hall of India
8. Sukumar Ghosh, Distributed Systems: An Algorithmic Approach, Second Edition
9. A.D. Kshemkalyani, M. Singhal, Distributed Computing: Principles, Algorithms, and Systems, ISBN: 9780521189842, Cambridge University Press, March 2011.

Introduction to Parallel algorithms by Jaja from Pearson, 1992.

Data Stream Management			
CourseCode:	**CD536**	**CourseCredits:**	**3**
CourseCategory:CC	**CC**	**Course(U/P)**	**U**
CourseYear(U/P):U	**3U**	**CourseSemester(U/P):**	**6U**
No.ofLectures+Tutorials(Hrs/Week):	**03+00**	**MidSem.ExamHours:**	**1.5**
TotalNo.of Lectures(L+T):30	**45+00**	**EndSem.ExamHours:**	**3**
COURSE OBJECTIVES			
1. To Understand Data mining principles and techniques.			
2. To Understand DM as a cutting edge business intelligence.			
3. To expose the students to the concepts of Data ware housing Architecture and Implementation.			
4. To study the overview of developing areas – Web mining, Text mining and ethical aspects of Data mining.			
5. To identify Business applications and Trends of Data mining.			
COURSEOUTCOMES			

At the end of the course the students should be able to:
1. Perform the preprocessing of data and apply mining techniques on it.
2. Identify the association rules, classification, and clusters in large data sets.
3. Solve real world problems in business and scientific information using data mining.
4. Use data analysis tools for scientific applications.
5. Implement various supervised machine learning algorithms.

UNIT I Introduction to data mining (DM)

Motivation for Data Mining - Data Mining-Definition and Functionalities – Classification of DM Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data Warehouse - Issues inDM – KDD Process

UNIT II Data Pre-processing

Data summarization, data cleaning, data integration and transformation, data reduction, data discretization and concept hierarchy generation, feature extraction , feature transformation, feature selection, introductionto Dimensionality Reduction, CUR decomposition

UNIT III Concept Description, Mining Frequent Patterns, Associations and Correlations

What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons, Basic concept, efficient and scalable frequent item-set mining methods, mining various kind of association rules, from association mining to correlation analysis, Advanced Association Rule Techniques, Measuring the Quality of Rules.

UNIT IV Classification and Prediction

Classification vs. prediction, Issues regarding classification and prediction, Statistical-Based Algorithms, Distance-Based Algorithms, Decision TreeBased Algorithms, Neural Network-Based Algorithms, Rule- Based Algorithms, Combining Techniques, accuracy and error measures, evaluation of the accuracy of a classifier or predictor. Neural Network Prediction methods: Linear and nonlinear regression, Logistic Regression Introduction of tools such as DB Miner / WEKA / DTREG DM Tools

UNIT V Cluster Analysis

Clustering: Problem Definition, Clustering Overview, Evaluation of Clustering Algorithms, Partitioning Clustering -K-Means Algorithm, KMeans Additional issues, PAM Algorithm; Hierarchical Clustering – Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering, Strengths andWeakness; Outlier Detection, Clustering high dimensional data, clustering Graph and Network data.

TextBooks:

1. Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, third edition 2011, ISBN: 1558604898.

2. Alex Berson and Stephen J. Smith, " Data Warehousing, Data Mining & OLAP",
Tata McGraw Hill

Predictive Analytics and Data Visualization			
CourseCode:	**CD-538**	**CourseCredits:**	**3**
CourseCategory: CC	**CC**	**Course(U/P)**	**U**
CourseYear(U/P):U	**3U**	**CourseSemester(U/P):**	**6U**
No.ofLectures+Tutorials(Hrs/Week):	**03+00**	**MidSem.ExamHours:**	**1.5**
TotalNo.of Lectures(L+T):30	**45+00**	**EndSem.ExamHours:**	**3**
COURSEOBJECTIVES			

1. To know regression methods
2. To know applications for model building prototyping and full scale the logical implications.
3. To know the implementation of forecasting of inventory models.
4. To know about managing resources, setting ticket prices.
5. To explore managing equipment maintenance, developing credit risk models.
COURSEOUTCOMES
Attheendofthecoursethestudentsshouldbeableto:
1. Financial services to aerospace.
2. Linear regression models and & least squares, multi regression.
3. Explain classification trees and boosting.
4. Explain reproducing kernels. SVM for classification
5. Numerical optimization, boosting methods.

UNIT I LINEAR METHODS OF REGRESSION AND CLASSIFICATION

Overview of supervised learning, Linear regression model and least squares, Multipleregression, Multiple outputs.

regression,Lassoregression,LinearDiscriminantAnalysis,Logisticregression,Perceptronlearning algorithm.

UNIT-II MODEL ASSESMENT AND SELECTION

Bias, Variance, andmodelcomplexity, Bias-variancetradeoff, Optimism of the training error rate, Estimate of In-sample prediction error, Effective number of parameters, Bayesian approach and BIC, Cross-validation, Boot strap methods, conditional or expected test error

UNIT-III ADDITIVE MODELS, TREES AND BOOSTING

Generalizedadditivemodels,Regressionandclassification trees ,Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting ,Examples (Spam data, California housing , NewZealand fish, Demographic data)

UNIT IV NEURAL NETWORK (NN) AND SUPPORT VECTOR

Introduction of Neural network and SupportVectorMachines(SVM),andK-nearestNeighbor: Fittingneural networks, Back propagation, Issues in training NN, SVM for classification,

Reproducing Kernels, SVM for regression, K-nearest –Neighbor classifiers(Image Scene Classification)

UNIT -V UNSUPERVISED LEARNING AND RANDOM FORESTS

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

TextBooks:

Trevor Hastie, Robert Tibshirani, Jerome Friedman, ***the Elements of Statistical Learning-DataMining,Inference,andPrediction,*** **SecondEdition, SpringerVerlag, 2009**

REFERENCES

1-**Annase Barrie**: Predictive Analytics for Dummies, 2013

Steven Finlay: Predictive Analytics and Data Mining 2014

Advanced Semantics Web Technology			
Course Code:	**CD-544**	**Course Credits:**	**3**
Course Category:	**CC9**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

UNIT I Introduction: 8 Introduction and Web Development Strategies, History of Web and Internet, Protocols governing Web, Writing Web Projects, Connecting to Internet, Introduction to Internet services and tools, Introduction to client-server computing. Core Java: Introduction, Operator, Data type, Variable, Arrays, Methods & Classes, Inheritance, Package and Interface, Exception Handling, Multithread programming, I/O, Java Applet, String handling, Event handling, Introduction to AWT, AWT controls, Layout managers.

UNIT II Web Page Designing: 8 HTML: list, table, images, frames, forms, CSS, Document type definition, XML: DTD, XML schemes, Object Models, presenting and using XML, Using XML Processors: DOM and SAX, Dynamic HTML.

UNIT III Scripting: 8 Java script: Introduction, documents, forms, statements, functions, objects; introduction to AJAX, VB Script, Introduction to Java Beans, Advantage, Properties, BDK, Introduction to EJB, Java Beans API.

UNIT IV Server Site Programming: 8 . Introduction to active server pages (ASP), Introduction to Java Server Page (JSP), JSP Application Design, JSP objects, Conditional Processing, Declaring variables and methods, Sharing data between JSP pages, Sharing Session and Application Data, Database Programming using JDBC, development of java beans in JSP, Introduction to Servelets, Lifecycle, JSDK, Servlet API, Servlet Packages, Introduction to COM/DCOM/CORBA

UNIT V PHP (Hypertext Preprocessor): 8 Introduction, syntax, variables, strings, operators, if-else, loop, switch, array, function, form, mail, file upload, session, error, exception, filter, PHP-ODBC,

Reference Books

Text books:

1. Burdman, Jessica, “Collaborative Web Development” Addison Wesley
2. Xavier, C, “ Web Technology and Design” , New Age International
3. Ivan Bayross,” HTML, DHTML, Java Script, Perl & CGI”, BPB Publication
4. Bhave, “Programming with Java”, Pearson Education
5. Herbert Schieldt, “The Complete Reference:Java”, TMH. 6. Hans Bergsten, “Java Server Pages”, SPD O’Reilly
6. Ullman, “PHP for the Web: Visual QuickStart Guide”, Pearson Education
7. Margaret Levine Young, “The Complete Reference Internet”, TMH

8. Naughton, Schildt, “The Complete Reference JAVA2”, TMH
9. Balagurusamy E, “Programming in JAVA”, TMH

Course Code:	**CS 625**	**Course Credits:**	**3**
Course Category:	**E4/DSE**	**Course (U / P)**	**U**
Course Year (U / P):	**3U**	**Course Semester (U / P):**	**6U**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
1. Understand the Big Data Platform and its Use cases
2. Provide an overview of Apache Hadoop
3. Provide HDFS Concepts and Interfacing with HDFS
4. Understand Map Reduce Jobs
5. Apply analytics on Structured, Unstructured Data. Exposure to Data Analytics with R.

COURSE OUTCOMES
At the end of the course the students should be able to:
1. Identify Big Data and its Business Implications Access and Process Data on Distributed File System
2. List the components of Hadoop and Hadoop Eco-System
3. Manage Job Execution in Hadoop Environment
4. Develop Big Data Solutions using Hadoop Eco System
5. Analyze Infosphere Big Insights Big Data Recommendations

UNIT I INTRODUCTION

Introduction to big data: Introduction to Big Data Platform, Challenges of Conventional Systems, Intelligent data analysis, Nature of Data, Analytic Processes and Tools, Analysis vs Reporting.

UNIT II DATA STREAMS

Mining data streams: Introduction to Streams Concepts, Stream Data Model and Architecture, Stream Computing, Sampling Data in a Stream, Filtering Streams, Counting Distinct Elements in a Stream, Estimating Moments, Counting Oneness in a Window, Decaying Window, Real time Analytics Platform (RTAP) Applications - Case Studies - Real Time Sentiment Analysis- Stock Market Predictions.

UNIT III HADOOP

Hadoop: History of Hadoop, the Hadoop Distributed File System, Components of Hadoop Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics, developing a Map Reduce Application-How Map Reduce Works, Anatomy of a Map Reduce Job run, Failures, Job Scheduling-Shuffle and Sort – Task execution, Map Reduce Types and Formats- Map Reduce Features Hadoop environment.

UNIT IV DATA PROCESSING

Frameworks: Applications on Big Data Using Pig and Hive, Data processing operators in Pig, Hive services, HiveQL, Querying Data in Hive, fundamentals of HBase and ZooKeeper, IBM InfoSphere Big Insights and Streams.

UNIT V DATA ANALYTICS TECHNIQUE

Predictive Analytics- Simple linear regression, Multiple linear regression., Interpretation 5 of regression coefficients. Visualizations, Visual data analysis techniques, interaction techniques, Systems and applications.

Text Books:

1. Michael Berthold, David J. Hand, “Intelligent Data Analysis”, Springer, 2007.
2. Tom White “Hadoop: The Definitive Guide” Third Edition, O’reilly Media, 2012.
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, “Understanding Big Data:
4. Analytics for Enterprise Class Hadoop and Streaming Data”, McGrawHill Publishing, 2012.

Data Science for Business Analytics			
Course Code:	**CD-542**	**Course Credits:**	**2**
Course Category:	**CC**	**Course (U / P)**	**U**
Course Year (U / P):	**1U**	**Course Semester (U / P):**	**1U**
No. of Lectures + Tutorials(Hrs./Week):	**02 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**30 + 00**	**End Sem. Exam Hours:**	**3**

COURSE OBJECTIVES
To make students understand and make inferences based on relations found in the sample, to relations in the population.
To Understand basic concepts in Excel
Design effective data visualizations in order to provide new insights into a research question or communicate information to the viewer.
Find and select appropriate data that can be used to create a visualization that answers aparticular research question.
For each individual statistical test students should be able to understand how it works, for what data and design it is appropriate and how results should be interpreted

COURSE OUTCOMES
At the end of the course the students should be able to:
Understand the fundamentals of Data Science and Excel
Key concepts in data science, including tools, approaches, and application scenarios
State-of-the-art tools to build data-science applications for different types of data, includingtext and CSV data
Work with frequency distribution, mean, covariance, serial correlation, multi-collinearity, conditional probability etc.
Analyse data using Sampling Distribution, t-distribution, F-distribution, Chi-Square distribution etc.

UNIT I Introduction.

Introduction to Data Science , Evolution of Data Science, Data Science Roles, Stages in a Data Science Project, Applications of Data Science in various fields, Data Security Issues. A description of the purpose

and application of Excel.

UNIT II Data Collection and Data Pre-Processing.

Data Collection Strategies, Data Pre-Processing Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization, Techniques for sorting and filtering data, including controlling the order of precedence in a sort, advanced filters.

UNIT III Exploratory Data Analytics.

Math's basic addition, subtraction, multiplication, division, multi brackets, Powers, Rounding, data set. Totals and counts – SUM(), COUNT() and COUNTA(). Other statistics. Conditional totals and counts. Advanced conditional sums, SUMPRODUCT and array formulas, Change the date system, format, or two digit year interpretation, Descriptive Statistics Mean, Standard Deviation, Cell referencing and naming, Creating named ranges, managing named ranges, Calculations with named ranges, and Automating processes with named ranges.

UNIT IV Model Development

Simple and Multiple Regression, Model Evaluation using Visualization, Residual Plot, Distribution Plot, Polynomial Regression and Pipelines, Measures for In-sample Evaluation, Prediction and Decision Making.

UNIT V Model Evaluation

Generalization Error, Out of Sample Evaluation Metrics, Cross Validation, Over fitting, Under Fitting and Model Selection, Prediction by using Ridge Regression, Testing Multiple Parametersby using Grid Search.

REFERENCE BOOKS:

1. Jojo Moolayil, "Smarter Decisions : The Intersection of IoT and Data Science",PACKT, 2016
.
2. Cathy O'Neil and Rachel Schutt , "Doing Data Science", O'Reilly, 2015.
3. David Dietrich, Barry Heller, Beibei Yang, "Data Science and Big data Analytics",EMC 2013

Block chain Technology for Data Science			
Course Code:	**CS-631**	**Course Credits:**	**3**
Course Category:	**CC9**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

UNIT I **Introduction to Blockchain**: Digital Money to Distributed Ledgers , Design Primitives: Protocols, Security, Consensus, Permissions, Privacy. Blockchain Architecture and Design: Basic crypto primitives: Hash, Signature,) Hashchain to Blockchain, Basic consensus mechanisms

UNIT II **Consensus:** Requirements for the consensus protocols, Proof of Work (PoW), Scalability aspects of Blockchain consensus protocols Permissioned Blockchains:Design goals, Consensus protocols for Permissioned Blockchains

UNIT III **Hyperledger** Fabric (A): Decomposing the consensus process , Hyperledger fabric components, Chaincode Design and Implementation Hyperledger Fabric (B): Beyond Chaincode: fabric SDK and Front End (b) Hyperledger composer tool

UNIT IV **Use case 1 :** Blockchain in Financial Software and Systems (FSS): (i) Settlements, (ii) KYC, (iii) Capital markets, (iv) Insurance
Use case 2: Blockchain in trade/supply chain: (i) Provenance of goods, visibility, trade/supply chain finance, invoice management discounting, etc

UNIT V **Use case 3:** Blockchain for Government: (i) Digital identity, land records and other kinds of record keeping between government entities, (ii) public distribution system social welfare systems Blockchain Cryptography, Privacy and Security on Blockchain

Reference Books

Text books:

1. Mstering Bitcoin: Unlocking Digital Cryptocurrencies, by Andreas Antonopoulos

2. Blockchain by Melanie Swa, O'Reilly

3. Hyperledger Fabric - https://www.hyperledger.org/projects/fabric

4. Zero to Blockchain - An IBM Redbooks course, by Bob Dill, David Smits - https://www.redbooks.ibm.com/
Redbooks.nsf/RedbookAbstracts.

Deep Learning in Biomedical and Genomics			
Course Code:	**CS-633**	**Course Credits:**	**3**
Course Category:	**CC9**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

UNIT I INTRODUCTION : Introduction to machine learning- Linear models (SVMs and Perceptrons, logistic regression)- Intro to Neural Nets: What a shallow network computes- Training a network: loss functions, back propagation and stochastic gradient descent- Neural networks as universal function approximates

UNIT II **DEEP NETWORKS : History of Deep Learning-** A Probabilistic Theory of Deep Learning Backpropagation and regularization, batch normalization- VC Dimension and Neural Nets-Deep Vs Shallow Networks-Convolutional Networks- Generative Adversarial Networks (GAN), Semisupervised Learning

UNIT III **DIMENTIONALITY REDUCTION** 9 Linear (PCA, LDA) and manifolds, metric learning - Auto encoders and dimensionality reduction in networks - Introduction to Convnet - Architectures – AlexNet, VGG, Inception, ResNet - Training a Convnet: weights initialization, batch normalization, hyper parameter optimization

UNIT IV **OPTIMIZATION AND GENERALIZATION :** Optimization in deep learning– Non-convex optimization for deep networks- Stochastic Optimization Generalization in neural networks- Spatial Transformer Networks- Recurrent networks, LSTM - Recurrent Neural Network Language Models- Word-Level RNNs & Deep Reinforcement Learning - Computational & Artificial Neuroscience

UNIT V CASE STUDY AND APPLICATIONS : Image net- Detection-Audio Wave Net- Natural Language Processing Word2Vec - Joint Detection-Bioinformatics- Face Recognition- Scene UnderstandingGathering Image Captions.

Reference Books

Text books:

1. Cosma Rohilla Shalizi, Advanced Data Analysis from an Elementary Point of View, 2015.
2. Deng & Yu, Deep Learning: Methods and Applications, Now Publishers, 2013.
3. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
4. Michael Nielsen, Neural Networks and Deep Learning, Determination Press, 2015.

INTERNET OF THINGS			
CourseCode:	**CS 639**	**CourseCredits:**	**3**
CourseCategory:CC	**CC**	**Course(U /P)**	**4U**
No.ofLectures+Tutorials(Hrs/Week):	**03 + 00**	**MidSem. ExamHours:**	**1.5**
TotalNo. ofLectures(L+T):30	**45**	**EndSem.ExamHours:**	**3**
COURSE OBJECTIVES			
1. Explore to the interconnection and integration of the physical world in IoT.			
2.Learning of networking concepts in IoT environment.			
3.Understanding of various wireless network, topologies, IoT protocols.			
4.Understad the importance of security issues in IoT.			
5.Implementation of IoT in real life with learning of tools like MATLAB.			
COURSE OUTCOMES			
At the end of the course the students should be able to:			
1 Figure out about all concepts of Internet of Things.			
2 understand building blocks of Internet of Things and its characteristics.			
3 learn application protocols for IoT.			
4 Able to understand the application areas of IoT.			
5 Realize the revolution of Internet in Mobile Devices, Cloud & Sensor Networks.			

UNIT I INTRODUCTION TO IOT

Genesis of IoT, IoT and Digitization, IoT Impact, Convergence of IT and OT, IoT Challenges, Drivers Behind New Network Architectures: Scale, Security, Constrained Devices and Networks, Data, Legacy Device Support .

UNIT II IOT NETWORK ARCHITECTURE AND DESIGN

Comparing IoT Architectures : The one M2M IoT Standardized Architecture, The IoT World Forum

(IoTWF) Standardized Architecture, Additional IoT Reference Models, A Simplified IoT Architecture, The Core IoT Functional Stack- Layer 1: Things: Sensors and Actuators Layer, Layer 2: Communications Network Layer, Layer 3: Applications and Analytics Layer, IoT Data Management and Compute Stack

:Fog Computing , Edge Computing, The Hierarchy of Edge, Fog, and Cloud.

UNIT III NETWORK AND APPLICATION PROTOCOLS FOR IOT

Wireless Communication Technologies: ZigBee, ESP8266, Introduction to sensors and modules - concept, layout, working, applications, Introduction of IoT Development Boards-Node MCU, Arduino, IoT Access Technologies 107IEEE 802.15.4, IEEE 802.15.4g and 802.15.4e, IEEE 1901.2a, IEEE

802.11ah, LoRaWAN, Constrained Devices, Constrained-Node Networks, Optimizing IP for IoT :From 6LoWPAN to 6Lo, Header Compression, Fragmentation, Mesh Addressing, Mesh-Under Versus Mesh- Over Routing, Authentication and Encryption on Constrained Nodes , Application Protocols for IoT: CoAP, Message Queuing Telemetry Transport (MQTT) .

UNIT IV DATA ANALYTICS AND SECURITY OF IOT

An Introduction to Data Analytics for IoT, Structured Versus Unstructured Data, Data in Motion Versus Data at Rest, IoT Data Analytics Overview, IoT Data Analytics Challenges, Machine Learning : Machine Learning Overview Supervised Learning, Unsupervised Learning, Neural Networks, Securing IoT : Common Challenges in IoT Security, Device Insecurity, Network Characteristics Impacting Security,

Security Priorities: Integrity, Availability, and Confidentiality, Formal Risk Analysis Structures: IAS OCTAVE, Top Vulnerabilities of Iot.

UNIT V. IMPLEMENTING IoT IN REAL LIFE

Interfacing sensors with development boards, communication modules with sensors, communication modules with development boards, MATLAB and Arduino Interfacing, Hands-on in IoT - various real life projects involving different boards, sensors, modules and communication technologies .

Text Books :

1. IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things by Rob Barton, Gonzalo Salgueiro, David Hanes
2. Vijay Madisetti and Arshdeep Bahga, “Internet of Things (A Hands-on-Approach)”, 1stEdition, VPT, 2014.
3. Francis daCosta, “Rethinking the Internet of Things: A Scalable Approach to Connecting Everything”,

1st Edition, Apress Publications, 2013

Data Analytics for Healthcare			
Course Code:	**CS-635**	**Course Credits:**	**3**
Course Category:	**CC9**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

UNIT I **Introduction to Data Analytics:** Sources and nature of data, classification of data (structured, semi-structured, unstructured), characteristics of data, introduction to Big Data platform, need of data analytics, evolution of analytic scalability, analytic process and tools, analysis vs reporting, modern data analytic tools, applications of data analytics. Data Analytics Lifecycle: Need, key roles for successful analytic projects, various phases of data analytics lifecycle – discovery, data preparation, model planning, model building, communicating results, operationalization.

UNIT II **Data Analysis:** Regression modeling, multivariate analysis, Bayesian modeling, inference and Bayesian networks, support vector and kernel methods, analysis of time series: linear systems analysis & nonlinear dynamics, rule induction, neural networks: learning and generalisation, competitive learning, principal component analysis and neural networks, fuzzy logic: extracting fuzzy models from data, fuzzy decision trees, stochastic search methods.

UNIT III **Mining Data Streams:** Introduction to streams concepts, stream data model and architecture, stream computing, sampling data in a stream, filtering streams, counting distinct elements in a

stream, estimating moments, counting oneness in a window, decaying window, Real-time Analytics Platform (RTAP) applications, Case studies – real time sentiment analysis, stock market predictions.

UNIT IV **Frequent Itemsets and Clustering:** Mining frequent itemsets, market based modelling, Apriori algorithm, handling large data sets in main memory, limited pass algorithm, counting frequent itemsets in a stream, clustering techniques: hierarchical, K-means, clustering high dimensional data, CLIQUE and ProCLUS, frequent pattern based clustering methods, clustering in non-euclidean space, clustering for streams and parallelism.

UNIT V **Frame Works and Visualization:** MapReduce, Hadoop, Pig, Hive, HBase, MapR, Sharding, NoSQL Databases, S3, Hadoop Distributed File Systems, Visualization: visual data analysis techniques, interaction techniques, systems and applications. Introduction to R - R graphical user interfaces, data import and export, attribute and data types, descriptive statistics, exploratory data analysis, visualization before analysis, analytics for unstructured data.

Reference Books

Text books and References:

1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
2. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press.
3. John Garrett,Data Analytics for IT Networks : Developing Innovative Use Cases, Pearson Education

INTERNET OF THINGS			
CourseCode:	**CS 639**	**CourseCredits:**	**3**
CourseCategory:CC	**CC**	**Course(U /P)**	**4U**
No.ofLectures+Tutorials(Hrs/Week):	**03 + 00**	**MidSem. ExamHours:**	**1.5**
TotalNo. ofLectures(L+T):30	**45**	**EndSem.ExamHours:**	**3**
COURSE OBJECTIVES			
1. Explore to the interconnection and integration of the physical world in IoT.			
2.Learning of networking concepts in IoT environment.			
3.Understanding of various wireless network, topologies, IoT protocols.			
4.Understad the importance of security issues in IoT.			
5.Implementation of IoT in real life with learning of tools like MATLAB.			
COURSE OUTCOMES			
At the end of the course the students should be able to:			
1 Figure out about all concepts of Internet of Things.			
2 understand building blocks of Internet of Things and its characteristics.			
3 learn application protocols for IoT.			
4 Able to understand the application areas of IoT.			
5 Realize the revolution of Internet in Mobile Devices, Cloud & Sensor Networks.			

UNIT I INTRODUCTION TO IOT

Genesis of IoT, IoT and Digitization, IoT Impact, Convergence of IT and OT, IoT Challenges, Drivers Behind New Network Architectures: Scale, Security, Constrained Devices and Networks, Data, Legacy Device Support .

UNIT II IOT NETWORK ARCHITECTURE AND DESIGN

Comparing IoT Architectures : The one M2M IoT Standardized Architecture, The IoT World Forum (IoTWF) Standardized Architecture, Additional IoT Reference Models, A Simplified IoT Architecture, The Core IoT Functional Stack- Layer 1: Things: Sensors and Actuators Layer, Layer 2: Communications Network Layer, Layer 3: Applications and Analytics Layer, IoT Data Management and Compute Stack

:Fog Computing , Edge Computing, The Hierarchy of Edge, Fog, and Cloud.

UNIT III NETWORK AND APPLICATION PROTOCOLS FOR IOT

Wireless Communication Technologies: ZigBee, ESP8266, Introduction to sensors and modules - concept, layout, working, applications, Introduction of IoT Development Boards-Node MCU, Arduino, IoT Access Technologies 107IEEE 802.15.4, IEEE 802.15.4g and 802.15.4e, IEEE 1901.2a, IEEE

802.11ah, LoRaWAN, Constrained Devices, Constrained-Node Networks, Optimizing IP for IoT :From 6LoWPAN to 6Lo, Header Compression, Fragmentation, Mesh Addressing, Mesh-Under Versus Mesh- Over Routing, Authentication and Encryption on Constrained Nodes , Application Protocols for IoT: CoAP, Message Queuing Telemetry Transport (MQTT) .

UNIT IV DATA ANALYTICS AND SECURITY OF IOT

An Introduction to Data Analytics for IoT, Structured Versus Unstructured Data, Data in Motion Versus Data at Rest, IoT Data Analytics Overview, IoT Data Analytics Challenges, Machine Learning : Machine Learning Overview Supervised Learning, Unsupervised Learning, Neural Networks, Securing IoT : Common Challenges in IoT Security, Device Insecurity, Network Characteristics Impacting Security,

Security Priorities: Integrity, Availability, and Confidentiality, Formal Risk Analysis Structures: IAS OCTAVE, Top Vulnerabilities of Iot.

UNIT V. IMPLEMENTING IoT IN REAL LIFE

Interfacing sensors with development boards, communication modules with sensors, communication modules with development boards, MATLAB and Arduino Interfacing, Hands-on in IoT - various real life projects involving different boards, sensors, modules and communication technologies .

Text Books :

4. IoT Fundamentals: Networking Technologies, Protocols, and Use Cases for the Internet of Things by Rob Barton, Gonzalo Salgueiro, David Hanes
5. Vijay Madisetti and Arshdeep Bahga, "Internet of Things (A Hands-on-Approach)", 1stEdition, VPT, 2014.
6. Francis daCosta, "Rethinking the Internet of Things: A Scalable Approach to Connecting Everything", 1st Edition, Apress Publications, 2013
7.

Modeling Process in Data Science			
Course Code:	**CD-530**	**Course Credits:**	**3**
Course Category:	**CC9**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

Unit 1- Introduction to modeling and simulation: Introduction to modeling, Examples of models, modeling of dynamic system, Introduction to simulation, MATLAB as a simulation tool, Bond graph modeling, causality, generation of system equations,

Unit 2- Bond graph modeling of dynamic system: Methods of drawing bond graph model- Mechanical systems & Electrical systems, some basic system modelsMechanical systems, Thermal systems, hydraulic systems, pneumatic systems and electrical systems.

Unit 3- System models of combined systems: Linearity and non linearity in systems combined rotary and translatory system, electro mechanical system, hydromechanical system,

Unit 4- Dynamic Response and System Transfer Function: Dynamic response of 1st order system and 2nd order system, performance measures for 2nd order system, system transfer function, transfer function of 1st and 2nd order system Block diagram algebra, signal flow diagram, state variable formulation, frequency response and bode plots.

Unit 5- Simulation and simulation applications: Simulation using SIMULINK, examples of simulation problems- simple and the compound pendulum, planner mechanisms, validation and verification of the simulation model, parameter estimation methods, system identifications, introduction to optimization,

Reference Books

Zeigler B.P. Praehofer. H. and Kim I.G. "Theory of modeling and simulation", 2nd Edition. Academic press
2000

- Robert L. Woods, Kent L. Lawrence, “Modeling and simulation of dynamic systems”, Person, 1997.
- Brown, Forbes T. “Engineering System Dynamics”, New York, NY: CRC, 2001. ISBN: 9780824706166.
- Pratab.R " Getting started with MATLAB" Oxford university Press 2009

Business Acumen and Intelligence			
Course Code:	**CD-540**	**Course Credits:**	**3**
Course Category:	**CC9**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

UNIT-1

BUSINESS INTELLIGENCE – INTRODUCTION: Introduction - History and Evolution: Effective and Timely decisions, Data Information and Knowledge, Architectural Representation, Role of mathematical Models, Real Time Business Intelligent System.

UNIT-2

BI – DATA MINING & WAREHOUSING: Data Mining - Introduction to Data Mining, Architecture of

Data Mining and How Data mining works(Process) , Functionalities & Classifications of Data Mining, Representation of Input Data, Analysis Methodologies. Data Warehousing - Introduction to Data Warehousing, Data Mart, Online Analytical Processing (OLAP) – Tools, Data Modelling, Difference between OLAP and OLTP, Schema – Star and Snowflake Schemas, ETL Process – Role of ETL.

UNIT-3

BI – DATA PREPARTTION: Data Validation - Introduction to Data Validation, Data Transformation – Standardization and Feature Extraction, Data Reduction – Sampling, Selection, PCA, Data Discretization.

UNIT-4

BI – DATA ANALYTICS PROCESS - Introduction to analytics process, Types of Analytical Techniques in BI –Descriptive, Predictive, Perspective, Social Media Analytics, Behavioral, Iris Datasets.

UNIT-5

IMPLEMENTATION OF BI – Business Activity Monitoring, Complex Event Processing, Business Process Management, Metadata, Root Cause Analysis.

Text Books:

1. Carlo-Vercellis, “Business Intelligence Data Mining and Optimization for Decision-Making”, First Edition
2. Drew Bentely, “Business Intelligence and Analytics” ,@2017 Library Pres., ISBN: 978-1-9789-2136-8
3. Larissa T. Moss & Shaku Atre, “Business Intelligence Roadmap: The Complete Project Lifecycle
4. For Decision-Support Applications”, First Edition, Addison-Wesley Professional,2003
5. Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., and Becker, B. John, “The Data Warehouse
6. Lifecycle Toolkit: Practical Techniques for Building Data Warehouse and Business Intelligence Systems”, Second Edition, Wiley & Sons, 2008.
7. Cindi Howson, “Successful Business Intelligence”, Second Edition, McGraw-Hill Education, 2013.

Regression for Time Series Analysis			
Course Code:	**CS-637**	**Course Credits:**	**3**
Course Category:	**CC9**	**Course (U / P)**	**P**
Course Year (U / P):	**1P**	**Course Semester (U / P):**	**1P**
No. of Lectures + Tutorials (Hrs/Week):	**03 + 00**	**Mid Sem. Exam Hours:**	**1.5**
Total No. of Lectures (L + T):	**45 + 00**	**End Sem. Exam Hours:**	**3**

UNIT-1

Introduction to R programming: History of R programming, starting and ending R, R as a scientific calculator , handling package, workspace, inspecting variables, operators and expressions in

R, data objects and types, vectors, matrices and arrays, lists and data frames, built-in and user-defined functions , strings and factors, flow control and loops, advanced looping, date and times.
Using R for statistical analysis: Importing data files, exporting data, outputting results, exporting graphs, graphics in R, interactively adding information of plot, performing data analysis tasks.
R commands for descriptive statistics, data aggregation, representation of multivariate data, code factorization and optimization, statistical libraries in R.

UNIT-2

Descriptive Statistics: Diagrammatic representation of data, measures of central tendency, measures of dispersion, measures of skewness and kurtosis, correlation, inference procedure for correlation coefficient, bivariate correlation, multiple correlations, linear regression and its inference procedure, multiple regression. Probability: Measures of probability, conditional probability, independent event, Bayes' theorem, random variable, discrete and continuous probability distributions, expectation and variance, markov inequality, chebyshev's inequality, central limit theorem.

UNIT-3

Inferential Statistics: Sampling & Confidence Interval, Inference & Significance. Estimation and Hypothesis Testing, Goodness of fit, Test of Independence, Permutations and Randomization Test, ttest/z-test (one sample, independent, paired), ANOVA, chi-square. Linear Methods for Regression Analysis: multiple regression analysis, orthogonalization by Householder transformations (QR); singular value decomposition (SVD); linear dimension reduction using principal component analysis (PCA).

UNIT-4

Pseudo-Random Numbers: Random number generation, Inverse-transform, acceptance-rejection, transformations, multivariate probability calculations. Monte Carlo Integration: Simulation and Monte Carlo integration, variance reduction, Monte Carlo hypothesis testing, antithetic variables/control variates, importance sampling, stratified sampling Markov chain Monte Carlo (McMC): Markov chains; Metropolis-Hastings algorithm; Gibbs sampling; convergence.

UNIT-5

Resampling Methods: Cross-validation, Bootstrapping, Jackknife resampling, percentile confidence intervals, permutation tests, Density Estimation: Univariate density estimation, kernel smoothing, multivariate density estimation Numerical Methods: Root finding; more on numerical integration; numerical maximization/minimization; constrained and unconstrained optimization; EM (ExpectationMaximization) algorithm; simplex algorithm.

References:

1. S.C. Gupta & V.K. Kapoor, "Fundamentals of Mathematical Statistics", Sultan Chand & Sons

2. Sheldon M. Ross, "Introduction to Probability and Statistics for Engineers and Scientists", Academic Press.

3. Dudewicz, E.J., Mishra, S.N., "Modern Mathematical Statistics", Willy

4. Purohit S. G., Gore S. D., Deshmukh S. K., "Statistics using R, Narosa

5. Rizzo, M. L., "Statistical Computing with R", Boca Raton, FL: Chapman & Hall/CRC Press

6. Normal Maltoff, The Art of R programming, William

7. Dalgaard, Peter, "Introductory statistics with R", Springer Science & Business Media

8. M. D. Ugarte, A. F. Militino, A. T. Arnholt, "Probability and Statistics with R", CRC Press

9. Kundu, D. and Basu, A., "Statistical computing – existing methods and recent developments", Narosa