SCHEME/SYLLABUS

For

Ph.D.COURSE

In

Computer Engineering

(For Session2020-21, Odd Semester)

DEPARTMENT OF COMPUTER ENGINEERING

FACULTY OF INFORMATICS & COMPUTING

J.C.BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY, YMCA

FARIDABAD

PhD (Computer Engineering)

Scheme of Studies / Examination (w.e.f. July, 2020)

Course No.	Course Title	Teaching Schedule			Marks For Sessionals	Marks for End Term Examination		TOTAL MARKS	CREDITS
		L	P	TOTAL		THEORY	PRACTICAL		
PHD-100A	**Research Methodology**	4	-	4	25	75	-	100	4
CPE-RPE	**Ethics**	4	-	4	25	75	-	100	4
	Elective-I	4	-	4	25	75	-	100	4
	TOTAL	12		12	75	225		300	12

Elective -1

1	Wireless Sensor Networks	PHDCE-17-01
2	Information Retrieval Systems	PHDCE-17-02
3	Advanced Information Retrieval Systems	PHDCE-17-03
4	Data Mining	PHDCE-18-01A
5	Advanced Internet of Things	PHDCE-18-01
6	Web Search and Information Retrieval	PHDCE-18-03
7	Machine Learning	MCS-18-106

8	Deep Learning	PHDCE-19-01
9	Digital Image Processing and Analysis	PHDCE-19-02
10	Big Data and Machine Learning Techniques	PHDCE-19-03
11	Agile Software development	PCE-02
12	Cloud Computing	MCS-18-304
13	Data Science	MCS-18-110
14	Advanced Software Testing	PHDCE-20-01
15	Computer Vision	MCS-18-208
16	Software Testing	MCS-18-209
17	Advances in Computer Vision and Image Processing	PHDCE-20-02
18	Security Aspects in Big Data	PHDCE-20-03

Note: Exam duration will be as under

(a)Theory exams will be of 3 hours duration

Code: PHD – 100A
RESEARCH METHODOLOGY
PhD (Common Subject)

No. of Credits: 4		Sessional:	25 Marks
L | T | P | Total		Theory:	75 Marks
4 | 0 | 0 | 4		Total:	100 Marks
		Duration of Exam:	3 Hours

Course Objectives:

- Understand research process in order to plan a research proposal
- Learn methods to devise and design a research set-up
- Plan and perform data collection methods and its analysis
- Conclude research in report writing

Course Outcomes: The research scholar shall be able to

CO1	Plan a research proposal and design the research.
CO2	Collect data through experiments or surveys as per research requirement.
CO3	Understand and apply sampling and sampling distributions.
CO4	Understand and perform quantitative and qualitative data analysis.
CO5	Write research report with proper citations.

Unit 1 **Introduction to Research**: Definition, need and purpose of research, types of research, research process, approaches to research, planning a research proposal, literature review.

Unit 2 **Measurement Scales**: Indexes vs. Scales, Types of Scale, construction of Scale, Bogardus social distance scale, Thurstone Scale, Likert Scale, Semantic Differential Scale, Guttmann Scale.

Unit 3 **Data Collection Methods:** Experiments and Surveys, Experiments: Classical Experiments, Independent & Dependent Variables, Pre Testing & Post Testing, Double Blind Experiment, Subject Selection, Variation on Experiment Design. Survey Research: Topics appropriate for survey research, Guidelines for asking questions, Questionnaire Construction, Strengths & Weakness of Survey Research, Types of Surveys.

Unit 4 **Sampling**: Types of sampling methods: Non Probability Sampling, Probability Sampling, Theory & Logic of Probability Sampling, Sampling Distributions & Estimates of Sampling Error.

Unit 5 **Data Analysis**: Qualitative v/s Quantitative data analysis, Qualitative Data Analysis: Discovering Patterns, Grounded Theory Method, Semiotics, Conversation Analysis, Qualitative Data Processing. Quantitative Data Analysis: Quantification of Data, Univariate Analysis, Bivariate Analysis, Multivariate Analysis, Regression Analysis, Description Analysis. Hypothesis. Multiple Attribute Decision Making.

Unit 6 **Report Writing, Ethical Issues and Outcomes**: Report Preparation, Structure of Report, Report Writing Skills, Citations, Research Papers, Intellectual Property Rights, Plagiarism, Patent, Commercialization, Ethical Issues.

References:

1. Research Methodology by R. Panneerselvam, 2nd Ed. PHI
2. Research Methodology by C.R. Kothari & Gaurav Garg, 3rd Ed. New Age Publishers
3. Research Methodology and Scientific Writing by C. George Thomas, Ane Books
4. The practice of social research by Earl Babbie, 14th Ed. Cengage
5. Multiple Attribute Decision Making, Gwo-Hshiung Tzeng and Jih-Jeng Huang, CRC Press

Research & Publication Ethics

Code: CPE-RPE

Phd- CommonSubject

Course structure

- The course comprises of six modules listed in table below. Each module has 4-5 units.

Modules	Unit title	Teaching hours
Theory		
RPE 01	Philosophy and Ethics	4
RPE 02	Scientific Conduct	4
RPE 03	Publication Ethics	7
Practice		
RPE 04	Open Access Publishing	4
RPE 05	Publication Misconduct	4
RPE 06	Databases and Research Metrics	7
	Total	**30**

Syllabus in detail

THEORY

- **RPE 01: PHILOSOPHY AND ETHICS (3 hrs.)**

1. Introduction to philosophy: definition, nature and scope, concept, branches
2. Ethics: definition, moral philosophy, nature of moral judgements and reactions

- **RPE 02: SCIENTIFICCONDUCT (5hrs.)**

1. Ethics with respect to science and research
2. Intellectual honesty and research integrity
3. Scientific misconducts: Falsification, Fabrication, and Plagiarism (FFP)
4. Redundant publications: duplicate and overlapping publications, salami slicing
5. Selective reporting and misrepresentation of data

- **RPE 03: PUBLICATION ETHICS (7 hrs.)**

1. Publication ethics: definition, introduction and importance
2. Best practices / standards setting initiatives and guidelines: COPE, WAME, etc.
3. Conflicts of interest
4. Publication misconduct: definition, concept, problems that lead to unethical behavior and vice versa, types
5. Violation of publication ethics, authorship and contributorship
6. Identification of publication misconduct, complaints and appeals
7. Predatory publishers and journals

PRACTICE

- **RPE 04: OPEN ACCESS PUBLISHING(4 hrs.)**

1. Open access publications and initiatives

2. SHERPA/RoMEO online resource to check publisher copyright & self-archiving policies
3. Software tool to identify predatory publications developed by SPPU
4. Journal finder / journal suggestion tools viz. JANE, Elsevier Journal Finder, Springer Journal Suggester, etc.

- **RPE 05: PUBLICATION MISCONDUCT (4hrs.)**

A. Group Discussions (2 hrs.)

1. Subject specific ethical issues, FFP, authorship
2. Conflicts of interest
3. Complaints and appeals: examples and fraud from India and abroad

B. Software tools (2 hrs.)

Use of plagiarism software like Turnitin, Urkund and other open source software tools

- **RPE 06: DATABASES AND RESEARCH METRICS (7hrs.)**

A. Databases (4 hrs.)

1. Indexing databases
2. Citation databases: Web of Science, Scopus, etc.

B. Research Metrics (3 hrs.)

1. Impact Factor of journal as per Journal Citation Report, SNIP, SJR, IPP, Cite Score
2. Metrics: h-index, g index, i10 index, altmetrics

****

Code: PHDCE-17-01

Subject Name: WIRELESS SENSOR NETWORKS

NO OF CREDITS: 4

	SESSIONAL:	**25**
L T P	**THEORY EXAM:**	**75**
4 0 0	**TOTAL :**	**100**

COURSE OBJECTIVES

1. To understand the basic WSN technology and its architecture.
2. To understand the medium access layer protocols and physical layer protocols.
3. To understand various routing challenges in network layer and study network layer protocols.
4. To study transport layer and Application layer protocols.
5. To study various challenges in security and types of security attacks in WSN.

Unit I : OVERVIEW OF ADHOC AND WIRELESS SENSOR NETWORKS

Wireless Networks, Infrastructure and Infrastructure less Wireless Networks, Ad hoc Wireless Networks, Types of Ad hoc Mobile Communications, Challenges Facing Ad hoc Mobile Networks, Issues in Designing a Routing Protocol for Ad Hoc Wireless Networks.

Basic definition and concepts of Wireless Networks, design principles for WSNs, comparison of MANET & WSN, Architecture and Protocol Stack of WSN, Unique constraints and challenges of WSNs, Applications of WSNs(Military Applications, Environmental Applications, Health Applications, Home Applications).

Unit II : MIDDLEWARE & OPERATING SYSTEMS FOR WSN WSN

Middleware Principles, Middleware Architecture, Existing Middleware (Milan, IrisNET,CLMF,MLM),Operating systems Design Issues, Examples of Operating Systems

Unit III : PHYSICAL & DATA LINK LAYER OF WSN

Physical Layer Technologies & Standards, Channel Coding,Source Coding, PHY Layer Standards, Challenges for MAC, Classification of MAC Protocols , Contention free and Contection Based MAC Protocols.

Unit IV : NETWORK LAYER

Challenges for Routing,Classification of Protocols, Data–Centric and Flat Architecture Protocols (Flooding,Gossiping,SPIN) Heirarchical protocols (LEACH,PEGASIS,TEEN,APTEEN), Location Based(Unicast, Multicast, GeoCast) and QoS based(Sequential Assignment,SPEED) Routing Protocols.

Unit V : TRANSPORT & APPLICATION LAYER

Challenges for Transport Layer, RMST Protocol, PSFQ Protocol, CODA Protocol, ESRT Protocol Source Coding, Query Processing, Network Management.

Unit VI : SECURITY IN WSN

Challenges of Security in WSN, Security Attacks in WSN, Protocols and Mechanisms for Security, IEEE 802.15.4 and ZigBee Security

COURSE OUTCOMES

After completing the course, the students should be able

1. To explain the basic concepts and applications of WSN.
2. To describe various protocols adopted in WSN.
3. To explain the architecture, features and performance of WSN.
4. To explain the various challenges faced while deploying a WSN.
5. To explain various standards adopted in WSN.

Code: PHDCE-17-02

Subject Name: INFORMATION RETRIEVAL SYSTEMS

NO OF CREDITS: 4

	SESSIONAL:	**25**
L T P	**THEORY EXAM:**	**75**
4 0 0	**TOTAL :**	**100**

Pre-requisites: Probability Theory, Database Management, Web Programming Course Objectives

Course Objectives

1. To build an understanding of the fundamental concepts of Information Retrieval.
2. To familiarize students with the basic taxonomy and terminology of Indices.
3. To learn about the Query Language and Metadata Search.
4. To learn about Semantic Web and various social networks.
5. To learn about the latest trends in Information Retrieval.

Unit 1: Introduction to Information Retrieval

Information retrieval problem, an inverted index, Processing Boolean queries, The extended Boolean model versus ranked retrieval, an inverted index, Bi-word indexes, Positional indexes, Combination schemes.

Unit 2: Index construction

Hardware basics, Blocked sort-based indexing, Single-pass in-memory indexing, Distributed indexing, Dynamic indexing, Other types of indexes Index compression: Statistical properties of terms in information retrieval, Heaps' law: Estimating the number of terms, Zipf's law: Modeling the distribution of terms, Dictionary compression, Dictionary as a string, Blocked storage, Postings file compression.

Unit 3: Scoring, term weighting and the vector space model

Parametric and zone indexes, Weighted zone scoring, Learning weights, The optimal weight, Term frequency and weighting, Inverse document frequency, Tf-idf weighting, The vector space model for scoring, Variant tf-idf functions.

Unit 4: Computing scores in a complete search system

Efficient scoring and ranking, Inexact top K document retrieval, Index elimination, Champion lists, Static quality scores and ordering, Impact ordering, Cluster pruning, Components of an information retrieval system, Tiered indexes.

Unit 5: Web search basics

Background and history, Web characteristics, The web graph, Spam, Advertising as the economic model, The search user experience, User query needs Crawling, Crawler architecture, DNS resolution, The URL frontier, Link analysis, The Web as a graph, Anchor text and the web graph, PageRank, Markov chains, The PageRank computation, Topic-specific PageRank

Unit 6: Language models for information retrieval
Language models, Finite automata and language models, Types of language models, Multinomial distributions over words, The query likelihood model, Using query likelihood language models in IR, Estimating the query generation probability, Language modelling versus other approaches in IR

Unit 7: Introduction to big data
Introduction to Big Data Platform, Challenges of Conventional Systems, Nature of Data, Analytic Processes and Tools.
Hadoop: History of Hadoop, the Hadoop Distributed File System, Components of Hadoop, Map Reduce-Type, Formats and Features.

COURSE OUTCOMES

After completion of course, students would be:

1. To identify basic theories and analysis tools as they apply to information retrieval.
2. To develop understanding of problems and potentials of current IR systems.
3. To learn and appreciate different retrieval algorithms and systems.
4. To apply various indexing, matching, organizing, and evaluating methods to IR problem.
5. To become aware of current experimental and theoretical IR research.

CODE: PHDCE-17-03

Subject Name: ADVANCED INFORMATION RETRIEVAL SYSTEMS

NO OF CREDITS: 4

	SESSIONAL: 25
L T P	**THEORY EXAM: 75**
4 0 0	**TOTAL : 100**

Pre-requisites: Probability Theory, Database Management, Web Programming, IRS

Course Objectives

1. To build an understanding of the fundamental concepts of Indices.
2. To familiarize students with the basic taxonomy and terminology of Semantic Web.
3. To learn about the basic elements of Semantic Web
4. To learn about real world Examples of Semantic Web and various social networks.
5. To learn about the design of semantic web.

Unit 1 : Introduction to Advanced Information Retrieval System
Information retrieval problem, an inverted index, Processing Boolean queries, The extended Boolean model versus ranked retrieval, Basic Index Construction & Index Compression- an inverted index, Bi-word indexes, Positional indexes, Combination schemes, Heaps Law & Zipf's law, Advanced Indexing & Retrieval Schemes, Better Link Analysis for ranking, Advantages of Advanced IRS over Basic IRS

Unit-2: The World Of Semantic Web
Introduction to WWW and definition of Search, Integration & Web Data Mining. Semantic Web and Introduction to metadata, its basic concepts and considerations. Search Engines for traditional web: Building the index table, conducting the search, Search Engines for Semantic Web: Building a semantic Web Search Engines, Using the Semantic Web Search Engines Problems related to Semantic Web Search Engines

Unit-3: The Basic Elements Of Semantic Web
Building block of Semantic Web: RDF, Overview of RDF, Basic elements of RDF, RDF Triples, Basic syntax, literal values and other RDF capabilities. Fundamental rules of RDF, Aggregation & Distributed Information RDFS, Taxanomy and Ontology, Core Elements of RDFS, Syntax & Examples, Concepts of Ontology & Taxonomy.

Unit-4: The Semantic Web: Real –World Examples
Swoogle: A Search Engine for Semantic Web documents.Swoogle Architecture, Discovery of SWDs, Collection of Metadata and calculation of rankings using metadata. Indexation and retrieval of SWDs FOAF: Friend of a Friend, Basic FOAF Vocabulary & Examples. Creating a FOAF Document and getting into the Circle. Semantic Markup, Semantic Markup issues.

Unit-5: Semantic Web Search Engines Revisited
Why Search Engines Again? Why traditional Search /engines fails? The Design of Semantic Web Search Engine prototype: Query processing, Discovery Strategy, Indexation Strategy, Using the prototype system.Why this prototype Search Engine provides better performance.

Course Outcomes

After completion of course, students would be:

1. To apply various indexing, matching, organizing, and evaluating methods to IR problem.
2. To develop understanding of problems and potentials Semantic Web.
3. To learn and appreciate basic elements of Semantic Web.
4. To become aware of current experimental and theoretical real world examples of Semantic Web.

CODE: PHDCE-18-01A

Subject Name: Data Mining

NO OF CREDITS: 4

	SESSIONAL:	**25**
L T P	**THEORY EXAM:**	**75**
4 0 0	**TOTAL :**	**100**

Pre-requisites: Database Management System

Course Objectives

1. To build an understanding of the fundamental concepts of datawarehousing.
2. To familiarize students with the basic taxonomy and terminology of Data Mining.
3. To learn about the basic rules of mining.
4. To learn about Clustering and Classification techniques, algorithms.
5. To learn about the recent trends and Web mining.

Unit 1: Introduction to Data Warehouse Data warehousing Definition, DBMS vs data warehouse, Three-tier architecture, Multidimensional data model, Various Schemas, OLAP operations, OLAP Servers, OLAP indexing, multi-feature cubes.

Unit 2: Introduction to Data Mining Data mining definition & task, KDD process, KDD versus data mining, data mining issues, data mining task primitives, supervised and unsupervised learning approaches, Data preprocessing.

Unit 3: Mining Association rules The a-priori algorithm and FP growth algorithm, generating rules, improving the efficiency of apriori, rule mining by partitioning, multi-dimensional and multi-level association rules, correlation rules; meta-rule guided mining and constraint based rule mining, Incremental rule mining.

Unit 4: Clustering techniques Cluster analysis, similarity and distance measures, partitioning methods: squared error, k-means and k-medoids approach; Hierarchical Clustering: agglomerative vs Divisive, Density based methods: Basic definitions and DBSCAN algorithm; Constraint based clustering.

Unit 5: Classification and Prediction Classification by Decision tree induction: information gain measure, Tree pruning methods, Bayesian classification, rule based classification, backpropagation through Neural Networks, Genetic Algorithm, Rough Sets, Support Vector Machines and Fuzzy techniques; Prediction: linear and non-linear regression techniques.

Unit 6: Recent trends and Web Mining Mining of Complex Data Objects, Spatial Databases, Temporal Databases; Web Mining, categories of web mining: web structure mining, web content mining and web usage mining, recent research in Data Mining and Web mining.

REFERENCES

1. Data Mining- Concepts & Techniques; Jiawei Han & Micheline Kamber- 2001, Morgan Kaufmann.

2. Data Mining: Introductory and advanced topics: Margaret H Dunham, S. Sridhar; Pearson education, 2008.

3. Data Warehousing in the Real World; Sam Anahory & Dennis Murray, Pearson.

4. Data Mining Techniques; Arun Pujar; 2001, University Press; Hyderbad.

5. Data Warehousing, Data Mining and OLTP; Alex Berson, 1997, Mc Graw Hill.

Course Outcomes

After completion of course, students would be able:

1. To become aware of basic concepts of data warehousing and data mining.
2. To develop understanding of basic rules of mining.
3. To learn and appreciate Clustering and Classification Techniques.
4. To become aware of recent trends and Web mining.

CODE: PHDCE-18-01

SUBJECT NAME: ADVANCED INTERNET OF THINGS

NO OF CREDITS: 4

			SESSIONAL:	25
L	T	P	THEORY EXAM:	75
4	0	0	TOTAL :	100

Pre-requisites: Internet and web Technology, Computer Networks

Course Objectives:

1. Student will be able to learn the basics of IOT.
2. Student will be able to analyse basic protocols of wireless and MAC.
3. Students will get basic knowledge of resource management.

MODULE-1: INTRODUCTION TO IoT

Introduction to IoT, Characteristics of IoT, Physical design of IoT, Logical design of IoT, Functional blocks of IoT, Communication models & APIs ,IoT & M2M Machine to Machine, Difference between IoT and M2M, Software define Network, Challenges in IoT(Design ,Development, Security).

MODULE-2: COMMUNICATION PROTOCOL AND SENSOR NETWORK

Wireless medium access issues, MAC protocol survey, Survey routing protocols, Protocols for machine to machine communication, Sensor deployment & Node discovery, Data aggregation & dissemination.

MODULE-3: RESOURCE MANAGEMENT IN IOT

Interoperability in IoT, Domain specific applications of IoT, Home automation, Industry applications, Surveillance applications, Other IoT applications Clustering, Synchronization, Software agents.

MODULE-4:

Recent trends in various learning techniques of machine learning and classification methods for IOT applications. Various models for IOT applications, Recent research paper studies.

Course Outcomes:

On successful completion of the course, the student will:

1. Understand the concepts of Internet of Things
2. Analyze basic protocols network
3. Design IoT applications in different domain and be able to analyze their performance

REFERENCES:

1. **Vijay Madisetti, Arshdeep Bahga, “Internet of Things: A Hands-On Approach”**
2. **Waltenegus Dargie,Christian Poellabauer, "Fundamentals of Wireless Sensor Networks: Theory and Practice"**

CODE: PHDCE-18-03

SUBJECT NAME: WEB SEARCH AND INFORMATION RETRIEVAL

NO OF CREDITS: 4

	SESSIONAL:	**25**
L T P	**THEORY EXAM:**	**75**
4 0 0	**TOTAL :**	**100**

Pre-requisites: Probability Theory, Database Management, Web Programming

Course Objectives:

1. To build an understanding of the fundamental concepts of Information Retrieval
2. To familiarize students with the basic taxonomy and terminology of Indices
3. To understand Heap‟s Law forestimation and Zipf‟s law for modeling distribution of terms 4. To understand dictionary compression and posting list compression
5. To introduce the scoring ,tf-idf weighting and vector space model for scoring
6. To understand cluster pruning and tiered indices
7. To learn the elements of Web Search basics
8. To learn various language models for information retrieval and their types

MODULE-1: INTRODUCTION TO INFORMATION RETRIEVAL

Information retrieval problem, an inverted index, Processing Boolean queries ,The extended Boolean model versus ranked retrieval , an inverted index ,Bi-word indexes, Positional indexes, Combination schemes

MODULE-2: INDEX CONSTRUCTION

Hardware basics, Blocked sort-based indexing , Single-passin-memory indexing ,Distributed indexing, Dynamic indexing, Other types of indexes Index compression: Statistical properties of terms in information retrieval ,Heap‟s law: Estimating the number of terms, Zipf’s law:Modeling the distribution of terms, Dictionary compression, Dictionarys a string, Blocked storage, Postings file compression.

MODULE-3: SCORING, TERM WEIGHTING AND THE VECTOR SPACE MODEL

Parametric and zone indexes ,Weighted zone scoring, Learning weights ,The optimal weight, Term frequency and weighting, Inverse document frequency, Tf-idf weighting, The vector space model for scoring, Variant tf-idf functions.

MODULE-4: COMPUTING SCORES IN A COMPLETE SEARCH SYSTEM

Efficient scoring and ranking, In exact top K document retrieval, Index elimination ,Champion lists, Static quality scores and ordering ,Impact ordering ,Cluster pruning ,Component so fan information retrieval system, Tiered indexes

MODULE-5: WEB SEARCH BASICS

Background and history, Web characteristics, The web graph, Spam, Advertising as the economic model, The search user experience, User query needs Crawling, Crawler architecture, DNS resolution, The URL frontier, Link analysis, The Web as a graph, Anchor text and the web graph ,Page Rank, Markov chains, The Page Rank computation, Topic-specific Page Rank

MODULE-6: LANGUAGE MODELS FOR INFORMATION RETRIEVAL

Language models, Finite automata and language models, Types of language models, Multinomial distributions over words , The query likelihood model, Using query like lihood language models in IR, Estimating the query generation probability ,Language modeling versus other approaches in IR

Course Outcomes:

a. The students will be able to understand basic Information Retrieval Systems.
b. The students will be able to lean how Boolean queries are processed.
c. The students will be able to identify the different types of indices: inverted index, positional index, bi-word index etc
d. The student will be able to make estimations and model distribution of terms and compressions
e. The students will be able to enumerate various types of indices. And also understand the concept of efficient storage of indices.
f. The students will be able to learn tf-idf scoring and vector space model scoring for ranking
g. The students will be able to understand Static quality ordering , cluster pruning and tiered indices h. The students will be able to understand the basic concept of Search Engines their architecture and various functional components.
i. The students will be able to understand the basic concept of Web crawlers and their architecture
j. The students will be able to understand various language models related to information retrieval

REFERENCES

1. C. D. Manning, P. Raghavan and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008 (available at http://nlp.stanford.edu/IR-book).
2. Chakrabarti, S. (2002). Mining the web: Mining the Web: Discovering knowledge from hypertext data. Morgan-kaufman.
3. B. Croft, D. Metzler, T. Strohman, Search Engines: Information Retrieval in Practice, AddisonWesley, 2009 (available at http://ciir.cs.umass.edu/irbook/).
4. R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley, 2011 (2nd Edition).
5. An Introduction to Information Retrieval Christopher D. Manning,Prabhakar Raghavan, HinrichSchütze Cambridge UniversityPress

CODE:MCS-18-106

SUBJECT NAME: MACHINE LEARNING

NO OF CREDITS: 4

	SESSIONAL:	**25**
L T P	**THEORY EXAM:**	**75**
4 0 0	**TOTAL :**	**100**

Course Objectives:

1. To learn the concept of how to learn patterns and concepts from data without being explicitly programmed in various IOT nodes.
2. To design and analyse 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.

MODULE-1: SUPERVISED LEARNING (REGRESSION/CLASSIFICATION)

- Basic methods: Distance-based methods, Nearest-Neighbours, Decision Trees, Naive Bayes
- Linear models: Linear Regression, Logistic Regression, Generalized LinearModels
- Support Vector Machines, Nonlinearity and Kernel Methods
- Beyond Binary Classification: Multi-class/Structured Outputs, Ranking

MODULE-2: 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)

MODULE-3:

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

MODULE-4:

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

MODULE-5:

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

MODULE-6:

Recent trends in various learning techniques of machine learning and classification methods for IOT applications, Various models for IOT applications.

Course Outcomes:

a. Extract features that can be used for a particular machine learning approach in various IOT applications.
b. 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.
c. To mathematically analyse various machine learning approaches and paradigms.

REFERENCES

1. Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Springer 2009 (freely available online)
3. Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2007.

CODE: PHDCE-19-01

SUBJECT NAME: DEEP LEARNING

NO OF CREDITS: 4

			SESSIONAL:	**25**
L	**T**	**P**	**THEORY EXAM:**	**75**
4	**0**	**0**	**TOTAL:**	**100**

NOTE: Question paper has two parts. Part-1 has 10 questions each of 2 marks. It covers the entire syllabus. Attempt any four questions out of six from Part-2.

Course Objectives:

1. To build an understanding of the fundamental concepts of Deep Learning
2. To familiarize students with the neural networks
3. To understand deep learning architecture
4. To introduce the concept of Classical Supervised Tasks with Deep Learning

Unit1: Introduction to Deep Learning, Bayesian Learning, Decision Surfaces

Unit2: Linear Classifiers, Linear Machines with Hinge Loss, Optimization Techniques, Gradient Descent, Batch Optimization

Unit 3: Introduction to Neural Network, Multilayer Perceptron, Back Propagation Learning, Unsupervised Learning with Deep Network, Auto encoders

Unit 4: Convolutional Neural Network, Building blocks of CNN, Transfer Learning, Revisiting Gradient Descent, Momentum Optimizer, Effective training in Deep Net- early stopping, Dropout, Batch Normalization, Instance Normalization, Group Normalization

Unit 5: Recent Trends in Deep Learning Architectures, Residual Network, Skip Connection Network, Fully Connected CNN etc.

Unit 6: Classical Supervised Tasks with Deep Learning, Image Denoising, Semanticd, Segmentation, Object Detection etc.

Course Outcomes:

a. The students will be able to understand deep learning concepts.
b. The students will be able to understand Linear Classifiers, Linear Machines with Hinge Loss, Optimization Techniques, Gradient Descent, Batch Optimization
c. The students will be able to understand Neural Network, Multilayer Perceptron, Back Propagation Learning, Unsupervised Learning with Deep Network, Auto encoders

The students will be able to understand Image Denoising, Semanticd, Semantic Web and related problems.

CODE: PHDCE-19-02

SUBJECT NAME: DIGITAL IMAGE PROCESSING AND ANALYSIS

NO OF CREDITS: 4

			SESSIONAL:	**25**
L	**T**	**P**	**THEORY EXAM:**	**75**
4	**0**	**0**	**TOTAL:**	**100**

Course Objectives: The aim of the course is to provide:

1. The basic principles and concepts in digital image processing ·
2. The application of digital image analysis moving towards image interpretation ·

Unit-1

Introduction to Digital Image Analysis, Characteristics of Digital Image Data, Elements of digital image processing and analysis systems: Digital image representation, visual perception, pixel connectivity, Digital Image Fundamentals: Image sensing and Acquisition; Image Sampling and Quantization.

Unit-2

Image Segmentation: Region-Based Segmentation; Thresholding, Edge detection, representation schemes, descriptors, Morphological Image processing, Object representation, description and recognition, Interpretation of Digital Image Data, Fourier transforms.

Unit-3

Image Processing Techniques, Visualization tools, Analysis tools and techniques, Colour image processing, Pixel relationship, Image enhancement, Smoothening, Sharpening, filtering, Compression and Restoration.

Unit-4

Study of Research papers on Digital Image Processing and Analysis.

Course Outcomes: Upon successful completion of the course students should be able to: ·

1. Describe the fundamental concepts and process flow of digital image analysis ·
2. Appropriately apply digital image analysis techniques to their research ·
3. Enhance their critical thinking skills .

TEXTBOOKS:

- Digital Image processing (3rd Edition) by Rafael C. Gonzalez, Richard E. Woods Publisher: Prentice Hall, ISBN-10: 013168728X, ISBN-13: 978-0131687288 [Available at SU's bookstore].
- Remote Sensing Digital Image Analysis (4th Edition) by John A. Richards and XiupingJia, Springer-Verlag Berlin Heidelberg.

REFERENCES:

- Anil Jain K. “Fundamentals of Digital Image Processing”, PHI Learning Pvt. Ltd., 2011.
- Willliam K Pratt, “Digital Image Processing”, John Willey, 2002.
- Malay K. Pakhira, “Digital Image Processing and Pattern Recognition”, First Edition, PHI Learning Pvt. Ltd., 2011.

CODE: PHDCE-19-03

SUBJECT NAME: BIG DATA AND MACHINE LEARNING TECHNIQUES
NO OF CREDITS: 4

			SESSIONAL:	**25**
L	**T**	**P**	**THEORY EXAM:**	**75**
4	**0**	**0**	**TOTAL:**	**100**

Course Objectives:

1. Understand big data for business intelligence. Learn business case studies for big data analytics.
2. Perform map-reduce analytics using Hadoop and related tools.
3. Understand machine learning algorithms.

MODULE-1:

What is big data, why big data, convergence of key trends, unstructured data, industry examples of big data, big data and marketing, fraud and bigdata, risk and big data, credit risk management, big data and algorithmic trading, big data and healthcare, big data in medicine, advertising and big data, big data technologies, open source technologies, cloud and bigdata, mobile business intelligence.

MODULE-2:

Introduction to Hadoop,, analyzing data with Hadoop, Hadoop streaming, design of Hadoop distributed file system (HDFS), HDFS concepts, MapReduce workflows, anatomy of MapReduce job run, classic Map-reduce, MapReduce types.

MODULE-3:

Hbase, data model and implementations, Hbase clients, Hbase examples, Cassandra, Cassandra data model, Cassandra examples, Cassandra clients, Hadoop integration, YARN, failures in classic Map-reduce and YARN, job scheduling, shuffle and sort, task execution.

MODULE 4

Introduction to machine learning, types of machine learning, Supervised learning: regression, classification, decision trees, naive bayes, Unsupervised learning: clustering: k-means/kernel k-means.

Course Outcomes:

After completion of course, students would be able to:

a. Describe big data and use cases from selected business domains.
b. Install, configure, and run Hadoop and HDFS.
c. Perform map-reduce analytics using Hadoop.
d. Use Hadoop related tools.
e. Familiar with Machine Learning.

REFERENCES

1. Michael Minelli, Michelle Chambers, and AmbigaDhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.
2. P. J. Sadalage and M. Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of
3. Polyglot Persistence", Addison-Wesley Professional, 2012.
4. Tom White, "Hadoop: The Definitive Guide", Third Edition, O'Reilley, 2012.
5. Eric Sammer, "Hadoop Operations", O'Reilley, 2012.
6. E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilley, 2012.
7. Lars George, "HBase: The Definitive Guide", O'Reilley, 2011.
8. Eben Hewitt, "Cassandra: The Definitive Guide", O'Reilley, 2010.
9. Alan Gates, "Programming Pig", O'Reilley, 2011.

CODE: MCS-18-304

SUBJECT NAME: CLOUD COMPUTING
NO OF CREDITS: 4

	SESSIONAL:	**25**
L T P	**THEORY EXAM:**	**75**
4 0 0	**TOTAL:**	**100**

NOTE: Question paper has two parts. Part-1 has 10 questions each of 2 marks. It covers the entire syllabus. Attempt any four questions out of six from Part-2.

Course Objectives:

1. The student will also learn how to apply trust-based security model to real-world security problems.
2. An overview of the concepts, processes, and best practices needed to successfully secure information within Cloud infrastructures.
3. Students will learn the basic Cloud types and delivery models and develop an understanding of the risk and compliance responsibilities and Challenges for each Cloud type and service delivery model.

MODULE-1:

Introduction to Cloud Computing

Online Social Networks and Applications, Cloud introduction and overview, Different clouds, Risks, Novel applications of cloud computing.

MODULE-2:

Cloud Computing Architecture

Requirements, Introduction Cloud computing architecture, On Demand Computing Virtualization at the infrastructure level, Security in Cloud computing environments, CPU Virtualization, A discussion on Hypervisors Storage Virtualization Cloud Computing Defined, The SPI Framework for Cloud Computing, The Traditional Software Model, The Cloud Services
Delivery Model.

Cloud Deployment Models

Key Drivers to Adopting the Cloud, The Impact of Cloud Computing on Users, Governance in the Cloud, Barriers to Cloud Computing Adoption in the Enterprise.

MODULE-3:

Security Issues in Cloud Computing

Infrastructure Security, Infrastructure Security: The Network Level, The Host Level, The Application Level, Data Security and Storage, Aspects of Data Security, Data Security Mitigation Provider Data and Its Security.

Identity and Access Management

Trust Boundaries and IAM, IAM Challenges, Relevant IAM Standards and Protocols for Cloud Services, IAM Practices in the Cloud, Cloud Authorization Management.

MODULE-4:

Security Management in the Cloud

Security Management Standards, Security Management in the Cloud, Availability Management: SaaS, PaaS, IaaS

Privacy Issues

Privacy Issues, Data Life Cycle, Key Privacy Concerns in the Cloud, Protecting Privacy, Changes to Privacy Risk Management and Compliance in Relation to Cloud Computing, Legal and Regulatory Implications, U.S.Laws and Regulations, International Laws and Regulations.

MODULE-5:
Audit and Compliance
Internal Policy Compliance, Governance, Risk, and Compliance (GRC), Regulatory/External Compliance, Cloud Security Alliance, Auditing the Cloud for Compliance, Security-as-a-Cloud.

MODULE-6:
ADVANCED TOPICS
Recent developments in hybrid cloud and cloud security.

Course Outcomes:
After completion of course, students would be able to:
a. Identify security aspects of each cloud model
b. Develop a risk-management strategy for moving to the Cloud
c. Implement a public cloud instance using a public cloud service provider
d. Apply trust-based security model to different layer

CODE: PCE-02

SUBJECT NAME: AGILE SOFTWARE DEVELOPMENT

NO OF CREDITS: 4

			SESSIONAL:	**25**
L	**T**	**P**	**THEORY EXAM:**	**75**
4	**0**	**0**	**TOTAL:**	**100**

NOTE: Question paper has two parts. Part-1 has 10 questions each of 2 marks. It covers the entire syllabus. Attempt any four questions out of six from Part-2.

Unit-1 Iterative and Evolutionary Development

Key motivations for iterative development, meeting the requirements and challenges iteratively, Risk driven & client-driven iterative planning, time-box iterative development, Evolutionary & adaptive development, Evolutionary requirement analysis, incremental & evolutionary delivery.

Unit-2 Adapting to Agile

Background, overview & definitions, Agile manifesto, Agile principles, Agile Methods-

Extreme Programming, Scrum Development methodology, Crystal family of methodologies, Rational Unified Process (RUP).

Unit-3 Agile Planning & Prioritization

Agile application to planning, Success features for Agile planning, velocity, prioritizing story & themes, Kano model for prioritization, Relative weighing model for prioritization.

Unit-4 Agile Estimation

Estimation, estimating size with story points, estimating with ideal days & ideal time, ideal day as a measure of time, techniques for estimating, re-estimation, choosing between story points & ideal days, Splitting user-stories, estimating user-stories, release plan, updating the release plan.

Unit-5 Agile Software Development using Scrum

Adaptive scrum, Product backlog, Sprint, Scrum life cycle, Scrum estimation, Scrum Planning, Working with scrum.

Unit-6 Agile Testing

Introduction to Agile testing quadrants, test-driven development, unit testing, component testing, functional testing, story testing, exploratory testing, scenario testing, usability testing, acceptance testing, performance and load testing, security testing, ility testing, pair testing,

References

1. **S/w development using Scrum, Succeeding with Agile by Mike Cohn**
2. **Agile & iterative development- A manager's Guide Craig Larman.**
3. **Agile Testing, A Practical Guide for Testers and Agile Teams,Lisa Crispin, Janet Gregory**

CODE: MCS-18-110

SUBJECT NAME: DATA SCIENCE

NO OF CREDITS: 4

			SESSIONAL:	**25**
L	**T**	**P**	**THEORY EXAM:**	**75**
4	**0**	**0**	**TOTAL:**	**100**

Course Objectives:

1. Provide you with the knowledge and expertise to become a proficient datascientist.
2. Demonstrate an understanding of statistics and machine learning concepts that arevital for datascience.
3. Produce Python/R code to statistically analyse adataset.
4. Critically evaluate data visualisations based on their design and usefor communicationgstories fromdata

MODULE-1:

Introduction to core concepts and technologies: Introduction, Terminology, datascience process, data science toolkit, Types of data, Example applications.

MODULE-2:

Data collection and management: Introduction, Sources of data, Data collection andAPIs, Exploring and fixing data, Data storage and management, Using multiple dataSources.

MODULE-3:

Data analysis: Introduction, Terminology and concepts, Introduction to statistics,Central tendencies and distributions, Variance,Distribution properties andarithmetic, Samples/CLT, Basic machine learning algorithms, Linear regression,SVM, NaiveBayes.

MODULE-4:

Data visualisation:Introduction, Types of data visualisation,Dataforvisualisation:Data types, Data encodings, Retinal variables, Mapping variables toencodings, Visual encodings.

MODULE-5:

Applications of Data Science,Technologies for visualization, Bokeh (Python)

MODULE-6:

Recent trends in various data collection and analysis techniques, variousvisualization techniques, application development methods of used in data science.

Course Outcomes:

After completion of course, students would be able to:

a. Explain how data is collected, managed and stored for datascience;

b. Understand the key concepts in data science, including their real-world applications and the

c. toolkit used by datascientists;

d. Implement data collection and management scripts usingMongoDB.

REFERENCES

1. Cathy O‟Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O‟Reilly.

2. Jsure Leskovek, AnandRajaraman and Jeffrey Ullman.Mining of Massive Datasets.v2.1,Cambridge UniversityPress.

CODE: PHDCE-20-01

SUBJECT NAME: ADVANCED SOFTWARE TESTING

NO OF CREDITS: 4

			SESSIONAL:	25
L	T	P	THEORY EXAM:	75
4	0	0	TOTAL :	100

Course Objectives

1. To get familiar the students about basic concepts of agile testing and its techniques.
2. To get familiar the students the concept of regression testing.
3. To study about the various testing automation and debugging tools and case studies.
4. To study the basic and advanced concepts of object-oriented testing

Unit 1 Agile Software Development and Agile Testing

Basics and Fundamentals of Agile Process Methods, Values of Agile, Principles of Agile, stakeholders, Challenges, Agile Testing: Agile Testing Techniques, Test-Driven Development, User Acceptance Test

Unit 2 Dynamic testing

White Box testing: Boundary value analysis, equivalence class portioning, state table based testing, decision table based, error guessing.

Black Box Testing: Logic coverage criteria, basic path testing, graph matrices.

Unit 3 Validation Testing

Unit testing, drivers, stubs, integration testing, methods, functional testing, system testing, recovery testing, security testing, stress testing, performance testing, usability testing

Unit 4 Regression Testing

Objective of regression testing, Regression test process, Regression testing techniques.

Unit 5 Test Automation and debugging

S/w measurement and testing, testing metrics and tools

Case Study: Testing for Object-oriented and web-based systems

Unit 6 Object-Oriented Testing

Use-case based testing; Class testing, Testing Exception handling

Unit 7 Data Science in Software Development

Big Data overview, Characteristics, Big Data Analytics, Application of Big Data- in science, manufacturing, health care, Government, Education, Information Technology.

Course Outcomes

a. The students will be able to understand the concepts of agile testing and its

techniques.

b. The Students will be able to implement Knowledge of verification and validation activities.

c. The Students will be able to apply black box and white box testing techniques.

d. The Students will be able to implement the concept of regression testing and its techniques.

REFERENCES

1. G.J Myers, The Art of Software Testing, John Wiley & Sons, 1979

2. Naresh Chauhan, Software Testing Principles and Practices,OXFORD University Press.

CODE: MCS-18-208
SUBJECT NAME: COMPUTER VISION

NO OF CREDITS: 4

			SESSIONAL:	25
L	T	P	THEORY EXAM:	75
4	0	0	TOTAL :	100

Pre-requisites: Linear algebra, vector calculus, Data structures and Programming.

Course Objectives:

1. Be familiar with both the theoretical and practical aspects of computing with images.
2. Have described the foundation of image formation, measurement, and analysis.
3. Understand the geometric relationships between 2D images and the 3D world.
4. Grasp the principles of state-of-the-art deep neural networks.

MODULE-1:

Overview, computer imaging systems, lenses, Image formation and sensing, Image analysis, pre-processing and Binary image analysis

MODULE-2:

Edge detection, Edge detection performance, Hough transform, corner detection.

MODULE-3:

Segmentation, Morphological filtering, Fourier transform.

MODULE-4:

Feature extraction, shape, histogram, color, spectral, texture, using CVIPtools,Feature analysis, feature vectors, distance /similarity measures, data preprocessing.

MODULE-5:Pattern Analysis

Clustering: K-Means, K-Medoids, Mixture of GaussiansClassification: Discriminant Function, Supervised, Un-supervised, Semisupervised Classifiers: Bayes, KNN, ANN models; Dimensionality Reduction: PCA, LDA,ICA, and Non- parametric methods.

MODULE-6:

Recent trends inActivity Recognition, computational photography, Biometrics.

Course Outcomes:

a. Developed the practical skills necessary to build computer vision applications.

b. To have gained exposure to object and scene recognition and categorization from images.

REFERENCES

1. Computer Vision: Algorithms and Applications by Richard Szeliski.

2. Deep Learning, by Goodfellow, Bengio, and Courville.

3. Dictionary of Computer Vision and Image Processing, by Fisher et al.

CODE: MCS-18-209

SUBJECT NAME: SOFTWARE TESTING

NO OF CREDITS: 4

	SESSIONAL:	25
L T P	THEORY EXAM:	75
4 0 0	TOTAL :	100

Pre-requisites: Knowledge of Software Engineering

Course Objectives

1. To get familiar the students about basic concepts of software testing and its techniques.

2. To study the concepts of Verification and validation activities.

3. To study in detail the process of performing the black box and white box testing approaches with examples.

4. To get familiar the students with the concept of regression testing, various testing automation and debugging tools and case studies.

5. To study the basic and advanced concepts of object oriented testing

MODULE-1: Testing terminology and Methodology

Definition of testing, goals, psychology ,model for testing, effective testing, limitations of testing, Importance of Testing, Definition of Failure, faults or bug, error, incident, test case, test ware, life cycle of bug, bug effects, bug classification, test case design, testing methodology, development of test strategy, verification, validation, Static testing: Inspection, Review and Walk through, dynamic testing ,testing life cycle model, testing techniques, testing principles, Testing Metrics.

MODULE -2: Verification and validation

Verification activities, verification of requirements, verification of HL design, verification of data design, verification of architectural design, verification of UI design, verification of LL design, introduction to validation activities

MODULE -3: Dynamic testing

White Box testing: Boundary value analysis, equivalence class portioning, state table based testing, decision table based, error guessing.

Black Box Testing: Logic coverage criteria, basic path testing, graph matrices.

MODULE -4: Validation Testing

Unit testing, drivers, stubs, integration testing, methods, functional testing, system testing, recovery testing, security testing, stress testing, performance testing, usability testing

MODULE -5: Regression Testing

Objectives of regression testing, Regression test process, Regression testing techniques.

MODULE -6: Test Automation and debugging

Software measurement and testing, testing metrics and tools.

Case Study: Testing for Object-oriented and web-based systems

MODULE -7: Object-Oriented Testing Use-case based testing; Class testing, Testing Exception handling

Course Outcomes:

a. The students will be able to understand the concepts of software testing, its techniques, verification andvalidation activities.

b. Study of black box, white box testing, regression testing and its techniques.

c. Study of object oriented testing techniques and testing

metrics.

d. Study of case studies and various testing automation and debugging tools.

REFERENCES

1. G.J Myers, The Art of Software Testing, John Wiley & Sons, 1979
2. Naresh Chauhan, Software Testing Principles and Practices, OXFORD University Press.

CODE: PHDCE-20-02
SUBJECT NAME: ADVANCES IN COMPUTER VISION & IMAGE PROCESSING

NO OF CREDITS: 4

			SESSIONAL:	25
L	T	P	THEORY EXAM:	75
4	0	0	TOTAL :	100

Pre-requisites: Basic co-ordinate geometry, matrix algebra, linear algebra & random process.

Course Objectives

1. To build an understanding of the fundamental concepts of Computer Vision.
2. To familiarize students with the basic taxonomy and terminology of Image Processing.
3. To learn about the features of an image &Image Segmentation.
4. To learn about basic Machine Learning techniques for Computer Vision.
5. To learn about the applications of computer vision.

Unit 1: Introduction to Computer Vision and Basic Concepts of Image Formation: Introduction and Goals of Computer Vision and Image Processing, Image Formation Concepts, Geometric Transformations, Geometric Camera Models, Camera Calibration, Image Formation in a Stereo Vision Setup, Image Reconstruction from a Series of Projections.

Unit 2: Image Processing Concepts: Image Transforms, Image Enhancement, Image Filtering, Colour Image Processing, Image Segmentation

Unit 3: Image Descriptors and Features: Texture Descriptors, Colour Features, Edges/Boundaries, Object Boundary and Shape Representations, Interest or Corner Point Detectors, Speeded up Robust Features, Saliency

Unit 4: Fundamentals of Machine Learning: Linear Regression, Basic Concepts of Decision Functions, Elementary Statistical Decision Theory, Parameter Estimation, Clustering for Knowledge Representation, Dimension Reduction, Linear Discriminant Analysis.

Unit 5: Applications of Computer Vision: Artificial Neural Network for Pattern Classification, Convolutional Neural Networks, Machine Learning Algorithms and their Applications in Image Segmentation, Motion Estimation and Object Tracking, Gesture Recognition, Face and Facial Expression Recognition, Image Fusion.

COURSE OUTCOMES

After completion of course, students would be able:

5. To understand the concepts of Image Pre-Processing& Computer Vision.

6. To develop understanding of challenges in computer vision.
7. To learn and appreciate different image segmentation techniques.
8. To understand the fundamentals of machine learning.
9. To become aware of applications and machine learning algorithms.

References:

1. MOOCs course by Prof. M. K. Bhuyan, “Computer Vision and Image Processing - Fundamentals and Applications”https://onlinecourses.nptel.ac.in/noc21_ee23/course
2. Forsyth & Ponce, “Computer Vision-A Modern Approach”, Pearson Education.
3. M.K. Bhuyan , “ Computer Vision and Image Processing: Fundamentals and Applications”, CRC Press, USA, ISBN 9780815370840 - CAT# K338147.
4. Richard Szeliski, “Computer Vision- Algorithms & Applications”, Springer.

*

CODE: PHDCE-20-03
SUBJECT NAME: SECURITY ASPECTS IN BIG DATA

NO OF CREDITS: 4

			SESSIONAL:	25
L	T	P	THEORY EXAM:	75
4	0	0	TOTAL :	100

COURSE OBJECTIVES:
This course is to help students

1. Outline the basic concepts and foundations of computer security,Identify concepts and ethics in Computer Security.
2. Demonstrate a breadth of knowledge in the topics of Computer Security, and understand its relevance and potential for an ever-increasing number of applications.
3. Possess the fundamental methodology for how to design and analyze security critical systems along with the identification of abnormalities caused by worms, viruses etc.
4. Learn and understand concepts of big data which include the study of modern computing big data technologies focusing on security.
5. To have knowledge about security-relevant decisions in context with 4 V's of Big data along with the working of programming tools like PIG, HIVE in Hadoop ecosystem.

UNIT – I Introduction to Security and Public key systems

Definition of Security, Traditional Cipher Systems, Stream ciphers, Block ciphers. Stream Ciphers, Cryptanalysis of stream ciphers, Block Ciphers: DES, modes of use of DES, AES, Diffie Hellman Key Exchange – Authentication and Digital signatures, Hash function – Authentication: Protocols – Digital Signature standards, RSA, MD-5, SHA, Firewall, types of Firewalls.

UNIT- II

Introduction to Big Data and security issues

Big Data - Introduction, Why Big Data, Types of Data, Characteristics of Big Data - The Four V's, advantages and disadvantages, Technology challenges for Big Data, Big Data Architecture, Applications of Big Data, Hadoop framework, Hadoop Distributed File System (HDFS), Comparison between HDFS and Google File system, Building Blocks of Hadoop, Introducing and configuring Hadoop cluster, MapReduce, Big Data Security issues, attacks and their solutions, Traditional Data Analytics v/s Big Data Analytics.

UNIT-III Big Privacy

Big Privacy: Content Privacy and Interaction Privacy, privacy preserving data publishing, Roles and operations of a privacy system, Privacy Research, Attacks: Active and Passive attacks, Linkage attack, probabilistic attack, Web Browsing attacks and Defence, Onion Routing, TOR System and Crowds System, Milestones of Privacy Study, Various privacy models, such as the k-anonymity, `diversity, t-

closeness, and ε-differential privacy, Disciplines in Privacy Study: Cryptography, Data Mining and Machine learning, Biometric Privacy, challenges of Privacy Study.

UNIT – IV Big Data Technologies

Introduction to Pig, Pig Architecture, Pig Latin Data model, operators, Introduction to Hive: Hive architecture, working with Hive Data Types, Creating and Managing Databases and Tables, views and indexes, Hive Data Manipulation Language, Querying and Analyzing Data, Introduction to Spark, Spark Architecture, Advantages over traditional data approaches, Spark Ecosystem, Spark for Big Data processing and its applications, Introduction to NoSQL, MongoDB.

COURSE OUTCOMES:

After completion of this course, students will be able to

1. Internalize the fundamental notions of threat, vulnerability, attack and countermeasure.
2. Categorize security threats, and the security services and mechanisms to counter them.
3. Understand Java concepts required for developing map reduce programs
4. Derive business benefit from unstructured data.
5. Learn the architectural concepts of Hadoop along with the Data Management Concepts in NoSQL environment.

Text / Reference Books:

1. Cryptography and Network Security, Atul Kahate, McGraw Hill.
2. Computer and Network Security, William Stallings, Pearson Publications.
3. Cryptography: Theory and Practice, Doughlas R Stinson, Maura B. Paterson, CRC Press, Taylor and Francis Group.
4. Hadoop: The Definitive Guide by Tom White, 3rd Edition, O'reilly
5. Hadoop in Practice by Alex Holmes, MANNING Publ.
6. Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization (2016), DT Editorial Services.