COURSES OF STUDIES

FOR PHDPROGRAM

(UNDER SEMESTER SYSTEM)

(Effective from January 2022)

GANGADHAR MEHER UNIVERSITY,
SAMBALPUR, ODISHA

PROGRAM OUTCOMES(POs):

PO1. **Knowledge and Understanding**: Develop an ability to understand the theoretical foundations of computer science for designing efficient methodologies along with the knowledge of limitations of computing.

PO2. **General, Technical and Professional Skills:** An ability to function effectively as an individual with diversified skills or as a part of a multi-disciplinary team setting to accomplish defined goals.

PO3. **Application of Knowledge and Skills**: Developing problem analysis skills and knowledge and applying the same in real life situation.

PO4. **Research Skills:** Explore research based knowledge and carry out academic investigations on the cutting edge technologies in allied subjects of Computer Science.

PO5. **General Learning Outcomes:** Create, select and apply advanced techniques and tools including modelling complex activities related to Computer Science.

PO6. **Constitutional, Humanistic, Ethical and Moral Values:** Design, develop and evaluate new system components or processes of computer science that meet the desired needs with appropriate considerations of industry, society, public health, safety, culture, environment and sustainable development sticking on to the ethics and values.

PO7: Employability Job Skills and Entrepreneur Skills: Prepare the students to take up a career as versatile contributors in industry, academia, research and development or entrepreneurship employing their expertise to advance personal growth while making meaningful contributions to societal progress

PROGRAM SPECIFIC OUTCOMES(PSOs):

PSO1. To shield students from the rapid obsolescence of computer technology, the program focuses on imparting foundational knowledge, fostering critical thinking skills, and cultivating technical expertise.

PSO2. Apply principles of computer science theory and concepts of software development to create effective computing-based solutions.

PSO3. Empowering the students to function as adept computer science professionals across various domains, including industry, advanced studies, research and development, academia, or entrepreneurship.

SEMESTER SYSTEM OFPh.D.

Sl No.	Paper	Semester	Credit	(Marks) Mid Term + Term End	Teaching Hours
1	711(Departmental)	Sem-1	4	100 (20+80)	48Hrs
2	712(Research Methgodology-1)	Sem-1	4	100 (20+80)	48Hrs
3	713(Research Methgodology-2)	Sem-1	4	100 (20+80)	48Hrs
4	714 Review Work	Sem-1	4	200(150+25+25)	48Hrs
TOTAL			20	500	160 Hrs

SEMESTER-1
Paper-711
Theory-Compulsory
Marks: 80+20 (Credits: 4)

List of Electives

711A	Data Warehousing and Mining
711B	Wireless Sensor Networks
711C	Internet of Things
711D	Machine Learning
711E	Information Security
711F	Soft computing
711G	Digital Image Processing
711H	Data Science

Course Code:	**711A**
Course Name:	**DATA WAREHOUSING AND MINING**
Category:	Programme Elective Course
Prerequisite:	Data Structure and Algorithm, Linear Algebra, Basics of Web programming

Paper – 711A

Data Warehousing and Mining

UNIT-I: **10hrs**

Evolution of Decision Support Systems- Data warehousing Components –Building a Data warehouse, Data Warehouse and DBMS, Data marts, Metadata, Multidimensional data model, OLAP vs OLTP, OLAP operations, Data cubes, Schemas for Multidimensional Database: Stars, Snowflkes and Fact constellations.

UNIT-II: **10hrs**

Types of OLAP servers, 3–Tier data warehouse architecture, distributed and virtual data warehouses. Data warehouse implementation, tuning and testing of data warehouse. Data Staging (ETL) Design and Development, data warehouse visualization, Data Warehouse Deployment, Maintenance, Growth, Business Intelligence Overview- Data Warehousing and Business Intelligence Trends - Business Applications.

UNIT-III: **10hrs**

Data mining-KDD versus datamining, Stages of the Data Mining Process-task premitives, Data Mining Techniques -Data mining knowledge representation – Data mining query languages, Integration of a Data Mining System with a Data Warehouse – Issues, Data preprocessing – Data cleaning, Data transformation, Feature selection, Dimensionality reduction, Discretization and generating concept hierarchies-Mining frequent patterns- association-correlation.
Decision Tree Induction - Bayesian Classifiation – Rule Based Classifiation –Classifiation by Back propagation – Support Vector Machines – Associative Classifiation – Lazy Learners –

Other Classifiation Methods.

UNIT-IV: **10hrs**

Clustering techniques – , Partitioning methods- k-means Hierarchical Methods - distance-based agglomerative and divisible clustering,
Mining complex data objects, Spatial databases, temporal databases, Multimedia databases, Time series and Sequence data; Text Mining –Graph mining-web mining-Application and trends in data mining

Text Books:

1. J. Han and M. Kamber, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann Publishers, 2011.
2. A. Berson and S. J. Smith, Data Warehousing, Data Mining & OLAP, 1st Edition, TataMc Graw Hill, 2007.
3. G. K. Gupta, Introduction to Data Min Data Mining with Case Studies, Easter Economy Edition, Prentice Hall of India, 2006.

Reference Books:

1. M. Kantardzic, Data mining concepts, models, methods, and algorithms, Wiley Inter science, 2003.
2. Witten, and E. Frank, Data Mining; Practical Machine Learning Tools and Techniques, third edition, Morgan Kaufmann, 2011.
3. G. M. Marakas, Modern Data Warehousing, Mining and Visualization, Prentice Hall, 2003

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to-
CO1	Describe the requirement of a data warehouse and its components.
CO2	Explain the data warehouse life cycle.
CO3	Explain the concepts of data mining and data pre-processing.
CO4	Analyze different classification algorithms and apply the same to real life problems.
CO5	Apply different clustering algorithms for solving problems in various domains.

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	**PSO3**
CO1	2						1	2	2	2
CO2	2						1	2	2	2
CO3	3		2	2			2	3	3	3
CO4	3		3	3	3		3	3	3	3
CO5	3		3	3	3		3	3	3	3

Course Code	**711B**
Course Name	**WIRELESS SENSOR NETWORKS**
Category	Programme Elective Course
Prerequisite	Basic Idea of Computer Networks

Paper –711 B

Wireless Sensor Networks

UNIT-I: **10hrs**

Introduction: Networked wireless sensor devices, Applications: Habitat Monitoring, Smart Transportation, Key design challenges. **Network deployment:** Structured versus randomized deployment, Network topology, Connectivity. Introduction to cloud system, Sensor Cloud Systems, Challenges in Sensor Cloud Systems.

UNIT-II: **08hrs**

Localization: issues & approaches, Coarse-grained & Fine-grained node localization, Network-wide localization. **Wireless characteristics:** Basics, Wireless link quality, Radio energy considerations, SINR capture model for interference.

UNIT-III: **10hrs**

Issues in designing MAC protocol for WSNs, Classification of MAC Protocols, Energy efficiency in MAC protocols, Asynchronous sleep techniques, Sleep-scheduled techniques.

Classification of Energy Management Schemes **Sleep-based topology control:** Constructing topologies for connectivity, constructing topologies for coverage.

UNIT-IV: **12hrs**

Routing: Metric-based approaches, Routing with diversity, Multi-path routing, Lifetime-maximizing energy-aware routing techniques, Geographic routing. Data-centric routing, Data-gathering with compression, Querying, Data-centric storage and retrieval, The database perspective on sensor networks.

Text Books:

1. Kazem Sohraby, Daniel Minoli, TaiebZnati, Wireless Sensor Networks: Technology, Protocols, and Applications, Wiley Inter Science.

2. Bhaskar Krismachari, Networking Wireless Sensors Cambridge University Press

Reference Books:

1. Edgar H. Callaway, Wireless Sensor Networks: Architectures and Protocols, Jr. Auerbach Publications, CRC Press.

2. Edited by C.S Raghavendra, Krishna M, Sivalingam, TaiebZnati, Wireless Sensor Networks, Springer.

3. Victor Lesser, Charles L. Ortiz, and Milind Tambe, Distributed Sensor Networks: A Multiagent Perspective, Kluwer Publications.

4.Feng Zhao, Leonidas Guibas , Morgan Kaufmann, Wireless Sensor Networks: An Information Processing Approach, Series in Networking 2004.

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to-
CO1	Define the basic concepts of wireless sensor networks, sensing, and challenges.
CO2	Explain various deployment structures of wireless sensor networks.
CO3	Describe and explore localization, radio standards and wireless characteristics.
CO4	Discuss the communication protocols adopted in wireless sensor networks and distinguish energy management schemes.
CO5	Analyze different routing techniques and identify various storage and retrieval issues.

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	**PSO3**
CO1	3							1		
CO2	3		1	1	1		1	1	1	1
CO3	3			1	1		1		1	
CO4	3		2	2	2		2	2	2	
CO5	3		3	3	3		3	2	2	3

Course Code	**711C**
Course Name	**INTERNET OF THINGS**
Category	Programme Elective Course
Prerequisite	Basic Idea of Computer Networks

DSE Paper –711C

Internet of Things

UNIT-I: 10hrs

Introduction to IoT, Basic requirements for building an IoT system, IoT reference framework, IoT network level – performance criteria.

IoT devices: Sensors, Types of sensors and their functions: temperature, pressure, air pollution, proximity, infrared, moisture & humidity, flow, level, noise, and speed sensors. Characteristics of sensors. Use of RFID

Actuators, Types of actuators and their functions: electrical, mechanical, and hydraulic actuators, controlling IoT devices

UNIT-II: 10hrs

IoT requirements for networking protocols, device addressing, credential management, wireless spectrum, determinism, security and privacy, application interoperability, semantic interoperability. IoT Protocol Stack: layered view.

Link layer: IEEE 802.15.4 technology, LoRaWAN end-to-end architecture, Time-Sensitive Networking

Internet Layer: Routing Protocol for Low-Power and Lossy Networks.

UNIT-III: 10hrs

Application Protocols Layer: Data Serialization Formats, Communication Paradigms: Request/Response Versus Publish/Subscribe, Blocking Versus Non-blocking, QoS: Resource Utilization, Data Timeliness, Data Availability, Data Delivery

IoT Application Protocols: CoAP, XMPP, MQTT, AMQP, SIP, IEEE 1888, and DDS RTPS.

Application Services Layer: ETSI M2M network architecture, oneM2M standards.

IoT Services Platform: Functions and Requirements, IoT Platform Manager, Discovery, Communication Manager, Data Management, Management of IoT Devices, Configuration and Fault Management, Performance Management and measures

UNIT-IV: 10hrs

IoT security and Privacy: challenges, requirements, IoT Three-Domain Architecture, Attacks and Countermeasures for each domain.

Applications of IoT in areas like Smart home, Agriculture, Healthcare, Industry, Transportation, Retail, Oil and Gas, Energy etc. IoT Service Model: Anything as a Service, IoT Connected Ecosystems Models

Text Book:

1. A. Rayes and S. Salam, Internet of Things from Hype To Reality: The Road to Digitization, 2nd Edition, Springer, 2019.

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to-
CO1	Describe basic concepts of IoT, its architecture and system design.
CO2	Employ the communication mechanisms between sensors and systems using various protocols and network models.
CO3	Explain IoT with respect to machine to machine and design IoT systems with data synchronization and resource manipulation. Explore various application protocols.
CO4	Discuss and describe different security issues and challenges.
CO5	Identify real world applications of IoT in multidisciplinary domains.

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	**PSO3**
CO1	3		2		1			1	3	3
CO2	3		3		3			1	3	3
CO3	3		3		3			1	3	3
CO4	3		3		3			1	3	3
CO5	3		3		3			1	3	3

Course Code:	**711D**
Course Name:	**MACHINE LEARNING**
Category:	Programme Elective Course
Prerequisite:	Basic knowledge of Mathematics

Paper-711D

Machine Learning

UNIT-I: **10hrs**

Introduction to machine Learning ((Supervised, Unsupervised and Reinforcement learning), Learning Models (Classification, Regression, Clustering).
Cluster Analysis, Partitioning Methods (k-Means, k-Medoids), Hierarchical Methods, Density-Based Methods, Evaluation of Clustering.

UNIT-II: **10hrs**

Conditional Probability, Bayes' Theorem, Naïve Bayes Classifier, K-nearest neighbour, Multiple linear regression, Shrinkage method, Ridge regression, Logistic regression, Linear Discriminant Analysis.

UNIT-III: **10hrs**

Neural Networks - Introduction, McP Neural Network, Perceptron Learning, Neural Networks - Backpropagation, Neural Networks - Initialization, Training & Validation. Decision Tree, Decision Tree Induction, Attribute Selection Measures, Information Gain, Gain Ratio, ID3, C4.5, Gini Index, CART.

UNIT-IV: **10hrs**

Support Vector Machine for linearly separable data, Kernel function, Support Vector Machine for linearly non-separable data.
Dimensionality reduction, Feature selection, Feature extraction, Principal Component Analysis.Model Cross- validation, Performance of Classification algorithms (Confusion Matrix, Precision and Recall).

Text Books:

1. T. Hastie, R. Tibshirani, and J. Friedman , The Elements of Statistical Learning-Data Mining, Inference, and Prediction, 2nd Edition, Springer Verlag, 2009.
2. S. Haykin, Neural Networks and Learning Machines, 3rdEdition,Pearson Education, 2009.

Reference Books:

1. Y. G. James, D. Witten, T. Hastie and R. Tibshirani, An introduction to Statistical learning with Applications in R, Springer, 2013.
2. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to-
CO1	Explain the concepts of supervised machine learning and its functionalities.
CO2	Perform classification using Bayes classifier, SVM, Decision Tree, and Random Forest.
CO3	Reduce dimension of feature space using feature selection and feature extraction.
CO4	Explain the concepts of unsupervised machine learning and its functionalities.
CO5	Apply supervised and unsupervised machine learning methods to solve real life problems.

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	**PSO3**
CO1	3		3		3			3	3	3
CO2	3		3		3			3	3	3
CO3	3		3		3			3	3	3
CO4	3		3		3			3	3	3
CO5	3		3		3			3	3	3

Course Code	**711E**
Course Name	**INFORMATION SECURITY**
Category	ProgrammeElective Course
Prerequisite	Computer Network

Paper-711E

Information Security

UNIT-I: **08hrs**

Attacks on Computers and Computer Security: Introduction, The need for security, Security goals, Security attacks(Attack on Confidentiality,Integrity,Availability)Security Services and Mechanisms, Techniques(Cryptography,Steganography).
Introduction to plain text and cipher text, encryption and decryption. substitution techniques, transposition techniques, symmetric and asymmetric key cryptography, steganography, possible types of cryptanalysis attacks.

UNIT-II: **12hrs**

Symmetric key Ciphers: Block Cipher principles &Algorithms(DES, AES, Blowfish), Differential and Linear Cryptanalysis, Block cipher modes of operation, Stream ciphers RC4,Location and placement of encryption function.
Introduction to number theory-Prime numbers,Euler's Phi-Function,Fermat's and Euler's theorem, Chinese Remainder Theorem,Generating Primes(MersennePrime,Fermat Prime),Primality testing(Deterministic algorithms,Probalistic algorithms)
Asymmetric key Ciphers: Principles of public key cryptosystems, Algorithms(RSA, Diffie-Hellman), Key Distribution.

UNIT-III: **10hrs**

Message Authentication Algorithms and Hash Functions: Message authentication (MDC,MAC)Nested MAC,HMAC,CMAC,Whirlpool. Hash functions: MD5 Message Digest algorithm,SHA-1. Digital signatures, Authentication Applications: Kerberos, X.509 Authentication Service, Public — Key Infrastructure, Biometric Authentication.

UNIT-IV: **10hrs**

E-Mail Security: Pretty Good Privacy, S/MIME IP **Security:lP** Security overview, IP Security architecture, Authentication Header, Encapsulating security payload, Combining security associations, key management.
Web Security: Web security considerations, Secure Socket Layer and Transport Layer Security, Secure electronic transaction. Intrusion Detection System(types, techniques).
Intruders, Virus and Firewalls: Intruders, Intrusion detection, password management, Virus and related threats, Countermeasures, Firewall design principles, Types of firewalls.

Text Book:

1. B. A. Forouzan,D.Mukhopadhyay,Cryptography and Network Security, 2nd Edition, McGraw Hill, 2007.
Reference Books:
1. A. Kahate, Network Security, 2nd Edition, McGraw Hill, 2008. 2. W. Stalling, Cryptography and Network Security, 4th Edition, Pearson Education, 2006.

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to-
CO1	Analyze the working of various Symmetric and Asymmetric key cryptographic algorithms for information security purpose
CO2	Identify the basic categories of threats in a networks
CO3	Able to demonstrate the design and use of hash functions, digital signatures, and key distribution with a wide range of key types
CO4	Discuss about Web security and Firewalls
CO5	Discuss about Intrusion Detection system.

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PS1**	**PS2**	**PS3**
CO1	2	2	2	1	3	1	1	2	2	1
CO2	2	1	2	1	2	1	1	2	1	2
CO3	2	2	2	2	2	2	2	2	2	2
CO4	2	2	2	1	2	1	2	2	2	1
CO5	2	2	2	2	2	2	2	2	2	2

Course Code	**711F**
Course Name	**SOFT COMPUTING**
Category	Programme Elective Course
Prerequisite	Basic Mathematics

Paper-711F

Soft Computing

UNIT-I: **08hrs**

Fuzzy Logic:Basic definition and terminology of fuzzy set, set theoretic operations, T-norm, T-conorm, Membership function formulation and parameterization, Extension Principle, Fuzzy relations, Linguistic variables, fuzzy if then rules, Compositional rule of inference, fuzzy reasoning, fuzzy inference systems, Mamdani fuzzy model.

UNIT-II: **08hrs**

Neural Network: Evolution of Artificial Neural Network (ANN), McP neural network, Perceptron, Perceptron convergence theorem, Perceptron learning, Multilayer perceptron, Back-propagation algorithm, Radial Basis Function, Radial Basis Function Network.

UNIT-III: **12hrs**

Genetic Algorithm: Introduction to Genetic Algorithm (GA), Working cycle of a GA, Binary coded GA, GA-parameter setting, Constraint handling, Advantages and disadvantages of GA.

UNIT-IV: **12hrs**

Simulated Annealing, Ant-Colony Optimization, Particle Swarm Optimization, Multi-objective optimization techniques and evolutionary computing.

Text Books:

1. J. Shing, R. Jang, C. T. Sun, and E. Mizutani, Neuro Fuzzy And Soft Computing - A Computational Approach to Learning and Machine Intelligence, 3rd Edition, Pearson Education, 2008.
2. D. K. Pratihar, Soft Computing, 2nd Edition, Narosa Publishing House, 2009.
3. S. Haykin, Neural Network - A Comprehensive Foundation, 2nd Edition, Pearson Education, 2006.

Reference Books:

1. T. Munakata, Fundamentals of the New Artificial Intelligence - Neural, Evolutionary, Fuzzy and More, 2nd Edition, Springer, 2014.
2. F. O. Karray and C. De Silva, Soft Computing and Intelligent System Design - Theory, Tools and Applications, 1st Edition, Pearson Education, 2009.

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to
CO1	Apply fuzzy logic and fuzzy inference system concept to design automation system for real life problems
CO2	Apply the concepts of genetic algorithm to solve engineering optimization problems.
CO3	Train the Artificial Neural Network for decision making in real life environment.
CO4	Use the concepts of Artificial Neural Network (ANN) to solve real life engineering and societal problems.
CO5	Apply the concepts of Simulated Annealing, Ant-Colony Optimization, Particle Swarm Optimization, Multi-objective optimization techniques to solve engineering optimization problems.

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	**PSO3**
CO1	3		3		3			3	3	3
CO2	3		3		3			3	3	3
CO3	3		3		3			3	3	3
CO4	3		3		3			3	3	3
CO5	3		3		3			3	3	3

Course Code	**711G**
Course Name	**DIGITAL IMAGE PROCESSING**
Category	Program Elective Course
Prerequisite	Basics of Digital Electronics and Basic understanding of calculus

Paper-711G

Digital Image Processing

UNIT-I: **08hrs**

Digital Image Fundamentals and Transforms: Elements of visual perception: Light, Brightness adaption and discrimination, Pixels, coordinate conventions, Imaging Geometry, Image sampling and quantization Basic relationship between pixels: Basic geometric transformations-Introduction to Fourier Transform and DFT : Properties of 2D Fourier Transform , FFT, Separable Image Transforms ,Walsh – Hadamard – Discrete Cosine Transform, Haar, Slant – Karhunen – Loevetransforms.Perspective Projection, Spatial Domain Filtering, sampling and quantization

UNIT-II: **08hrs**

Image Enhancement Techniques: Spatial Domain methods: Basic grey level transformation, Histogram equalization, Image subtraction, Image averaging, Spatial filtering: Smoothing, sharpening filters,Laplacian filters, Frequency domain filters : Smoothing, Sharpening filters,Homomorphic filtering.

UNIT-III: **16hrs**

Image Restoration and Image Compression: Model of Image Degradation/restoration process: Noise models, inverse filtering, least mean square filtering, constrained least mean square filtering, blind image restoration, Pseudo inverse, Singular value decomposition.
Lossless compression: Variable length coding: LZW coding, Bit plane coding- predictive coding, DPCM.
Lossy Compression: Transform coding – Wavelet coding – Basics of Image compression standards: JPEG, MPEG, Basics of Vector quantization

UNIT-IV: **08hrs**

Image Segmentation and Representation: Edge detection: Thresholding, Region Based segmentation, Boundary representation: chair codes, Polygonal approximation, Boundary segments: boundary descriptors: Simple descriptors, Fourier descriptors, Regional descriptors, Simple descriptors, Texture.

Text Book:

1. Digital Image Processing, 3rd Edition, by Rafael C Gonzalez and Richard E Woods. Publisher: Pearson Education.

Reference Books:

1. Fundamentals of Digital Image Processing, By Anil K Jain
2. Digital Image Processing, By William K Pratt, John Willey (2001)
3. Image Processing Analysis and Machine Vision, By MillmanSonka, Vaclav hlavac, Roger Boyle, Broos/colic, Thompson Learniy (1999).
4. Digital Image Processing and Applications, By, B. Chanda, D. DuttaMagundar, Prentice Hall of India, 2000

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to-
CO1	Understand the need for image transforms different types of imagetransforms and their properties
CO2	Develop any image processing application.
CO3	Understand the rapid advances in Machine vision
CO4	Learn different techniques employed for the enhancement of images
CO5	Understand a digital image and differentprocessing techniques for the better analysis of an image

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	**PSO3**
CO1	3		3		3			3	3	3
CO2	3		3		3			3	3	3
CO3	3		3		3			3	3	3
CO4	3		3		3			3	3	3
CO5	3		3		3			3	3	3

Course Code	**711H**
Course Name	**Data Science**
Category	Programme Elective Course
Prerequisite	Statistics, Mathematics, Programming Knowledge

Paper-711H

Data Science

UNIT-I: **08hrs**

Brief Introduction to Data Science. Descriptive statistics, notion of probability, distributions, mean, variance, covariance, covariance matrix, hypothesis testing

UNIT-II: **08hrs**

Introduction to Machine Learning: Supervised Learning, Decision Tree Induction, Naïve Bayes Classification, Rule based Classification, K-Nearest Neighbour, Unsupervised Machine learning, Clustering, K-Means, Association rule mining, Apriori, FP-Tree.

UNIT-III: **12hrs**

Attribute-oriented analysis: Attribute generalization, Attribute relevance, Class comparison, Statistical measures, Data pre-processing: Data cleaning, Data transformation, Data reduction. Predictive Modelling: Regression, Decision Tree, SVM.

UNIT-IV: **12hrs**

Feature selection (Filters; Wrappers), Dimensionality reduction: PCA and LDA.
Ensemble Learning, Bagging, Boosting, Gradient Boosting (Random Forest, Adaptive Boosting)
Time Series Data Analysis: Introduction to Time Series, Correlation, Forecasting (Univariate): Autoregressive Moving Average (ARMA) models, Autoregressive Integrated Moving Average (ARIMA) models, Introduction to Deep Learning.

Text Books:

1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly. 2014.

2. James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning with applications in R. Springer, 2013.Joel Grus, Data Science from Scratch: First Principles with Python. 1st Edition.

3. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson.

4. Laura Igual and Santi Seguí, Introduction to Data Science, Springer

Reference Books:

1.Han, J., Kamber, M., Pei, J. Data mining concepts and techniques. Morgan Kaufmann, 2011.
2. "Practical Data Science with R". Nina Zumel, John Mount. Manning, 2014.
3. Davy Cielin, Arno Meysman, Mohamed Ali, Introducing Data Science, Manning
4. Andreas, Practical Data Science, Apress

	COURSE OUTCOMES: After completion of this course successfully, the students will be able to-
CO1	Develop in depth understanding of the key technologies in data science and business analytics: data mining, machine learning, visualization techniques, predictive modelling, and statistics.
CO2	Practice problem analysis and decision-making.
CO3	Gain practical, hands-on experience with statistical programming languages and tools through coursework and applied research experiences.
CO4	Analyze and interpret data using an ethically responsible approach.
CO5	Apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively.

Mapping of COs to POs (1: Low, 2: Medium, 3: High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PSO1**	**PSO2**	**PSO3**
CO1	3		3		3			3	3	3
CO2	3		3		3			3	3	3
CO3	3		3		3			3	3	3
CO4	3		3		3			3	3	3
CO5	3		3		3			3	3	3

PAPER-712
RESEARCH METHODOLOGY-I

Unit-I

SCOPE, PHILOSOPHY AND ETHICS OF RESEARCH AND ETHICS

i) Introduction and Scope

ii) Introduction to philosophy: definition, nature and scope, concept, branches

iii) Ethics: definition, moral philosophy, nature of moral judgments and reactions, Research ethics, Institutional ethics committee.

iv) Ethics with respect to science and research

v) Intellectual honesty and research integrity

Unit-II

SCIENTIFIC CONDUCT

i) Research problem: Identification, Selection, Formulation of research objectives

ii) Research design: Components, Types and Importance

iii) Scientific misconducts: Falsification, Fabrication, and Plagiarism (FFP)

iv) Redundant publications: duplicate and overlapping publications, salami slicing

v) Selective reporting and misrepresentation of data

Unit-III

TECHNICAL WRITING

i) Literature search technique, using SCOPUS, Google Scholar, PUBMED, Web of science, Indian Citation Index, and RG

ii) Types of technical documents; Full length research paper, Short/Brief communications, Letters to editor, Book chapter, Review, Conference report, Project proposal Components of a full length research paper; , Rationale of the paper, Aims and objectives, Hypothesis building, Work plan, Materials and methodology, Results and discussion, Conflict of interest statement,

iii) Components of a research proposal; Project summary Key words, Origin of the proposal, Major objectives Methodology, Instrument facility available in the PI's department, Overview of status of Research and Development in the subject, Importance of the proposed project in the context of current status.

iv) Styles of referencing; APA, MLA, Oxford, Harvard, Chicago, Annotated bibliography, Tools for citing and referencing, Grammarly, Endnote etc, How to cite and how to do referencing

Unit-IV

PUBLICATION ETHICS

i) Publication ethics: definition, introduction and importance

ii) Best practices / standards setting initiatives and guidelines: COPE, WAME, etc.
iii) Conflicts of interest
iv) Publication misconduct: definition, concept, problems that lead to unethical behavior and vice versa, types
v) Violation of publication ethics, Subject specific ethical issues, FFP, authorship,
vi) Identification of publication misconduct, complaints and appeals
vii) Predatory publishers and journals
viii) Plagiarism-Pitfall
ix) Use of plagiarism software like Turnitin, Urkund and other open source software tools, . Complaints and appeals: examples and fraud from India and abroad

PAPER-713
RESEARCH METHODOLOGY-II

Unit-I

IPR AND CYBER LAW.

i) Patents, Patent laws, process of patenting a research finding
ii) Intellectual property (IP), Intellectual property right (IPR)
iii) Copyright, Trademarks, GI
iv) Cyber laws
v) COPE

Unit-II

QUANTITATIVE DATA ANALYSIS

i) Types of Data, Data Collection – Methods and Tools
ii) Hypothesis testing
iii) Normal and Binomial distributions and their property
iv) Tests of significance: Student *t*- test, *F*- test, *Chi-square* test
v) Correlation and Regression
vi) ANOVA – One-way and Two-way, Multiple-range test

Unit-III

COMPUTER FUNDAMENTALS

i) Introduction to MS-Office software: MS-Word(Track change)
ii) MS-Excel
iii) MS-Power Point
iv) Features for Statistical Data Analysis Tool Pack, SPSS
v) Tables, Figures and Pictures using Excel
vi) Preparation of Posters
vii) Electronic submission of manuscripts
viii) Communication skills, oral and poster

Unit-IV

ADVANCED TOOLS & TECHNIQUES IN RESEARCH

i) Indexing databases
ii) Citation databases: Web of Science, Scopus, etc.
iii) Research Metrics
iv) Impact Factor of journal as per Journal Citation Report, SNIP, SJR, IPP, Cite Score
v) Metrics: h-index, g index, i10 index, altmetrics
vi) Open access publications and initiatives
vii) SHERPA/RoMEO online resource to check publisher copyright & self-archiving policies

viii) Software tool to identify predatory publications developed by SPPU

ix) Journal finder /journal suggestion tools viz. JANE, Elsevier Journal Finder, Springer Journal Suggester, etc.

Paper-714
REVIEW WORK
Each Ph.D (Computer Science) student is required compulsorily to select a problem on any area of Computer Science, carry out intensive review of literature and prepare a report to be presented followed by viva-voce. The evaluation of review work shall be jointly made by both the external and internal examiners. The marks shall be distributed as follows:- Written report- 150 Presentation - 25 Viva-Voce – 25. **End Term-200**