Course Code: PPHD-CS-131
Course Title: Research Methodology.
Credits: 4

Maximum Marks : 100
University Examination: 100
Duration of Examination:3 Hours

Objective:
The course will acquaint the scholars how to obtain information, compile and analyze the same, and presentation of results in a research assignment.

Unit-I

Research: Definition and meaning of Research, Characteristics of research, Types of Research. Steps in research (Identification, Selection and Formulation of research problem), Research Question, Research Design, Formulation of Hypothesis, Review of Literature.
Books: General books, Text books, Reference books.
Electronic Source: Born Digital & Digitized; online & offline; open Access, Subscribed; Digital Divide; Technophobia & Technophilia; Information haves & have-nots.

Unit-II

Sampling Techniques: Sampling theory, types of sampling, Steps in Sampling, Sampling and Non-sampling error, Sample size, Advantages and limitations of sampling. Collection of Data: Primary Data, Meaning, Data Collection methods, Secondary Data, Meaning, Relevance, Limitations and Cautions.

Unit-III

Statistics in Research: Measure of central tendency, Dispersion, Hypothesis, Fundamentals of Hypothesis Testing, Standard Error, Point and Interval estimates. Important Non-Parametric tests: Sign, Run, Kruskal.

Unit-IV

Para metric Tests: Testing of significance, mean, proportion, Variance and Correlation, testing for Significance between means, proportions, Variances and correlation co-efficient. Chi-square tests, ANOVA-One-way and Two-way.

Unit-V

Research Report: Types of reports, contents, styles of reporting, Steps in drafting reports, Editing the final draft, Evaluating the final draft.

Note for Paper Setting:
The question paper shall comprise of 10 long answer type questions , two questions from each unit and each question will carry 20 marks and the candidates will be required to answer five questions selecting one from each unit.

References:

1. **S.P. Gupta**, " Statistical Methods"
2. **C.R. Kothari**, " Research Methodology Methods and Techniques".
3. **B.N Gupta**, " Statistics (Theory and Practice)".
4. **Santosh Gupta**, "Research Methodology Methods and Techniques".

Course Code: PPHD-CS-132
Course Title: Research and Publication Ethics (RPE).
Credits: 2

Maximum Marks : 50
University Examination: 50
Duration of Examination:2 Hours

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

References

Bird, A. (2006). *Philosophy of Science*. Routledge.

MacIntyre, Alasdair (1967) *A Short History of Ethics*. London.

P. Chaddah, (2018) Ethics in Competitive Research: Do not get scooped; do not get plagiarized, ISBN:978-9387480865

National Academy of Sciences, National Academy of Engineering and Institute of Medicine. (2009). *On Being a Scientist: A Guide to Responsible Conduct in Research: Third Edition*. National Academies Press.

Resnik, D. B. (2011). What is ethics in research & why is it important. *National Institute of Environmental Health Sciences*, 1–10. Retrieved from https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm

Beall, J. (2012). Predatory publishers are corrupting open access. Nature, 489(7415), 179–179. https://doi.org/10.1038/489179a

Indian National Science Academy (INSA), Ethics in Science Education, Research and Governance(2019), ISBN:978-81-939482-1-7. http://www.insaindia.res.in/pdf/Ethics_Book.pdf

Course Code: PPHD-CS-EL-211
Course Title: Principles of Machine Learning
Credits: 4

Maximum Marks : 100
University Examination: 100
Duration of Examination:3 Hours

Objective
The course is designed to introduce the Machine Learning concepts and its learning. The syllabus is applied in nature and is recommended for those who want to pursue research in the said area.

Unit I

Machine Learning: Introduction to Machine Learning, Paradigms of Machine Learning, Supervised Learning, Unsupervised Learning, Reinforcement Learning. Preprocessing of Data, Choosing a Machine Learning Model.

Unit II

Supervised Learning: Classification Algorithms: Naive Bayesian, multinomial and Binomial, Decision Tree, Random Forest, Linear Discriminant Analysis, Fischer Linear Discriminant Analysis

Unit III

Regression: Linear Regression, Logistic Regression, K-Nearest Neighbor Classification Basis, Support Vector Machine, Linear SVM, Non-Linear SVM.

Unit IV

Unsupervised Learning: Clustering Algorithms, K-Means Algorithms, Hierarchical Clustering, Frequent Pattern Mining Algorithms

Unit V

Python: Introduction, Data Types, Functions, modules and Packages, Loading packages and namespaces, reading and writing data, Control Flow, Applications of machine Learning using Pathon.

Note for Paper Setting:

The question paper shall comprise of 10 long answer type questions, two questions from each unit and each question will carry 20 marks and the candidates will be required to answer five questions selecting one from each unit.

Textbook:

1. **M. Tom**, " Machine Learning", **McGraw-Hill**, 1997 First Edition.
2. **M. Alex, R. Anna, H. Steve**, "Python in a Nutshell", **Oreilly Publications**, 5th Edition

References:

1. **A. Ethem**, "Introduction to Machine Learning", Edition 2nd , **MIT Press** .

Course Code: PPHD-CS-EL-212
Course Title: Big Data Analytics
Credits: 4

Maximum Marks : 100
University Examination: 100
Duration of Examination:3 Hours

Objective:

The course is designed to introduce the Big Data and its allied tools for analyzing the data. The syllabus is applied in nature and is recommended for those who want to pursue research in the said area.

Unit I

Data Mining: Introduction to Data mining, Data Mining Vs Traditional Analysis, Knowledge Discovery in Databases (KDD), Data mining tasks, Application areas of Data mining, Types of data in Data mining, Issues and challenges in Data mining.

Unit II

Data pre-processing: Introduction to Data Pre-processing, Data Cleaning: Missing Values, Noisy Data & Data cleaning as a process. Data Integration, Data Transformation, Data Reduction: Data Cube Aggregation, Attribute subset selection & Dimensionality reduction.

Unit III

Data Mining Tasks: Clustering: Cluster Analysis and its applications, Clustering Methods (Partitioning, Hierarchical, Density and Grid based methods). Classification: Classification and its utility in Data Mining, Classification methods (Probabilistic and Decision Tree). Association Rule Mining: Association rule mining and its utility in Data mining, Fundamentals of Association rule mining, Apriori Algorithm for Association rule mining.

Unit IV

Big Data: Evolution of Big data, Importance of Big data, Characteristics of Big data (5 V's of Big data), Data Analytics and its categories (Descriptive, Diagnostic, Predictive and Perspective analytics), Different types of data in Big data: Structured, Semi-structured and Unstructured type, Big data analytics life cycle, challenges of Big data.

Unit V

Technologies for Handling Big data: Introduction to Hadoop, RDBMS Vs Hadoop, Functioning of Hadoop, Hadoop Distributed File System and its architecture, Map-Reduce architecture, Hadoop Ecosystem.

Note for Paper Setting:

The question paper shall comprise of 10 long answer type questions, two questions from each unit and each question will carry 20 marks and the candidates will be required to answer five questions selecting one from each unit.

Textbook:

1. M. J. Zaki and W. Meira Jr., **Data Mining and Analysis – Fundamental Concepts and Algorithms**, Cambridge Press.
2. J. Han and M. Kamber, **Data Mining Concepts and Techniques**, 2nd Ed., Morgan Kaufman.
3. Minelli M., Chambers M., Dhiraj A., **Big Data, Big Analytics: Emerging Business.**

Course Code: PPHD-CS-EL-212
Course Title: Big Data Analytics
Credits: 4

Maximum Marks : 100
University Examination: 100
Duration of Examination:3 Hours

Objective:
The course is designed to introduce the Big Data and its allied tools for analyzing the data. The syllabus is applied in nature and is recommended for those who want to pursue research in the said area.

Unit I

Data Mining: Introduction to Data mining, Data Mining Vs Traditional Analysis, Knowledge Discovery in Databases (KDD), Data mining tasks, Application areas of Data mining, Types of data in Data mining, Issues and challenges in Data mining.

Unit II

Data pre-processing: Introduction to Data Pre-processing, Data Cleaning: Missing Values, Noisy Data & Data cleaning as a process. Data Integration, Data Transformation, Data Reduction: Data Cube Aggregation, Attribute subset selection & Dimensionality reduction.

Unit III

Data Mining Tasks: Clustering: Cluster Analysis and its applications, Clustering Methods (Partitioning, Hierarchical, Density and Grid based methods). Classification: Classification and its utility in Data Mining, Classification methods (Probabilistic and Decision Tree). Association Rule Mining: Association rule mining and its utility in Data mining, Fundamentals of Association rule mining, Apriori Algorithm for Association rule mining.

Unit IV

Big Data: Evolution of Big data, Importance of Big data, Characteristics of Big data (5 V's of Big data), Data Analytics and its categories (Descriptive, Diagnostic, Predictive and Perspective analytics), Different types of data in Big data: Structured, Semi-structured and Unstructured type, Big data analytics life cycle, challenges of Big data.

Unit V

Technologies for Handling Big data: Introduction to Hadoop, RDBMS Vs Hadoop, Functioning of Hadoop, Hadoop Distributed File System and its architecture, Map-Reduce architecture, Hadoop Ecosystem.

Note for Paper Setting:

The question paper shall comprise of 10 long answer type questions, two questions from each unit and each question will carry 20 marks and the candidates will be required to answer five questions selecting one from each unit.

Textbook:

1. M. J. Zaki and W. Meira Jr., **Data Mining and Analysis – Fundamental Concepts and Algorithms**, Cambridge Press.
2. J. Han and M. Kamber, **Data Mining Concepts and Techniques**, 2nd Ed., Morgan Kaufman.
3. Minelli M., Chambers M., Dhiraj A., **Big Data, Big Analytics: Emerging Business.**