Course work for PhD students aspiring to work in "Algorithms and AI-ML"

The following courses will be offered to the first semester Ph.D. students aspiring to work in the area of Algorithms and AI-ML:

- P.Py: Programming in Python
- M.Ml: Mathematics for Machine Learning
- ML: Machine Learning
- DAA: Design and Analysis of Algorithms
- RM: Research Methodology

The time table and details of each courses is given below.

Routine

	10:00 - 11:30	11:45 – 13:15	14:15 – 15:45	16:00 – 17:30
Monday	-	-	DAA	P.Py
Tuesday	-	RM	-	ML
Wednesday	M.Ml	-	RM	-
Thursday	-	-	-	P.Py
Friday	M.Ml	-	DAA	ML

1. Programming in Python

A. **Description:**
This course provides a platform to learn basic programming concepts. It also helps to learn how to handle and pre-process real-life data in ML and NLP domains.

B. **Pre-requisites:** None

C. **Outline of the syllabus**

1. **Introduction**
   I. Programming Languages with a managed run-time
   II. Installation, Compilation, and Execution
   III. Python IDE's (Notebooks, Editors)
2. **Basic**
   I. Indentation
   II. Data types: Number, List, Dict. String, tuples etc.
   III. Loop: for, iterations
   IV. Built-in functions: for, String, List, etc.
   V. Input / Output
3. **Object Oriented Features**
4. **Functional Programming**
5. **Learning about Libraries and Packages**
   I. Python Libraries and Packages
   II. Decorators
   III. Error Handling and Debugging
6. **Scripting, Scraping, Pattern Matching, Plotting**
   I. Beautifulsoup, regex, Matplotlib
7. **Natural Language Processing with Python**
   I. Nltk, spacy, scipy, etc.
8. **Machine Learning Basics with Python**
   I. Numpy, Pandas, Tensorflow, Pytorch, etc
9. **Additional: Web Development with Python**

D. **Duration** 45 hours (15 weeks, 3 hours per week).

E. **Learning outcome and the objective of the course**
Students will be proficient in programming in Python and will be able to develop projects to complement their research in Machine Learning, Data science and other areas in Computer Science.

F. **Books and References**

(a) Introduction to Machine Learning with Python
Andreas C. Müller, Sarah Guido
First Edition, O'Reilly Media, Inc.

(b) Python Machine Learning By Example
Yuxi (Hayden) Liu
Third Edition, Packt Publishing

G. **Assessment methodology**
Written and Programming assignments, Examinations.

H. **Pedagogic methodology**
Lectures, presentations, and Programming session.

The syllabus for this course is prepared by Ayan Bandopadhaya.

2. Mathematics for Machine Learning

A. **Description**
This course will introduce the basic mathematics and statistics required for the understanding of the Machine Learning models and tools. The students should intuitively connect an underlying mathematical principles and its application in addressing a specific ML task.

B. **Pre-requisites**
None. But it will be helpful if the students are eager to understand the intuition of a specific formulation.

C. **Outline of the syllabus**

1. **Basic Probability:** Probability Theory: Combinatorics, probability rules and axioms, Bayes theorem
2. **Random variable and distributions:** random variable, variance, expectation, conditional and joint distribution, Bernoulli, Binomial, Multinomial, uniform, Gaussian distribution
3. **MLE, MAP:** Maximum-likelihood-estimation, prior-posterior, MAP, sampling methods.
4. **Stochastic and Markov Processes:** Basic ideas, properties and applications.
5. **Matrices:** Transpose, Inverse, Ranks, Definiteness and semi-definiteness, eigenvalue eigenvectors
6. **Applications:** PCA, SVD, LU Decomposition Assignment Problem.
7. **Vector Spaces:** Knowledge required to intuitively understand the ML basics.
8. **Applications of calculus:** Basic concepts of Maxima-minima, Lagrange multiplier, Hessian matrix. Topics may be added according to the time availability and the mutual enthusiasm of the instructor and the scholars.

D. **Duration** 45 hours (15 weeks, 3 hours per week).

E. **Learning outcome and the objective of the course**
To have an intuitive understanding of the mathematical principles, which can aid the understanding of the real-world problems and ML-based solutions.

F. **Books and References**

(a) **Mathematics for Machine Learning**: Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong

(b) **Basic Probability and Theory**: Robert Ash.

G. **Assessment methodology**
Written examination (Mid-sem + end-sem) and few assignments.

H. **Pedagogic methodology**
Lectures, illustrations, (offline mode of learning is preferred).

The syllabus for this course is prepared by Payel Sadhukhan.

3. Machine Learning

A. **Description**
An introduction to Machine Learning will be provided, with topics covering Supervised, Unsupervised and Semi-Supervised Learning, and an introduction to Reinforcement Learning and Machine Learning for Time Series data. Recent topics of interest such as Interpretable and Explainable AI and Ethics in AI will also be covered. The Machine Learning models covered in this course will be applied on data from different domains such as Computer Vision, Speech and Language processing, Bioinformatics, etc.

B. **Pre-requisites**
Undergraduate level Calculus and Linear Algebra.

C. **Outline of the syllabus**

1. **Regression:** Linear Regression, Model Fitting, Regression Splines, Generalized Additive Models (9 hours).
2. **Classification:** Bayes Classifier, k-NN Classification, Logistic Regression, PAC Learning, Support Vector Machines, Vapnik-Chervonenkis Dimensions, Multi-Layered Perceptrons (12 hours).
3. **Tree Based Learning:** Decision Trees, Random Forests, XGBoost (6 hours).
4. **Dimension Reduction:** Principal Component Analysis, Subspace / Low Rank Estimation, Non-linear Subspace Projections (6 hours).
5. **Clustering:** k-Means, Hierarchical, Density-Based, Mean Shift, Subspace and Spectral Clustering (9 hours).
6. **Semi-Supervised Learning** (6 hours).
7. **Reinforcement Learning** (6 hours)
8. **Time-Series Data and Continuous Learning** (3 hours)
9. **Selected Topics in ML:** Interpretable and Explainable AI, Ethics in AI (3 hours).

D. **Duration**
Duration in weeks: 20 weeks
Hours per week: 3 hours

E. **Learning outcome and the objective of the course**
Understanding the mathematical motivations behind different machine learning methods. Developing machine learning methods and applying them to different kinds of data, followed by analysing the outcomes.

F. **Books and References**

(a) Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani: An Introduction to Statistical Learning with Applications in R, Second Edition. Springer, 2021.
(b) Trevor Hastie, Robert Tibshirani, and Jerome Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer, 2017.
(c) Ulisses Braga-Neto: Fundamentals of Pattern Recognition and Machine Learning. Springer, 2020.
(d) Richard O. Duda, Peter E. Hart, and David G. Stork: Pattern Classification, Second Edition. Wiley, 2000.
(e) Christopher M. Bishop: Pattern Recognition and Machine Learning. Springer, 2006.

G. **Assessment methodology**
Written and programming assignments, exams.

H. **Pedagogic methodology**
Lectures, presentations, and programming sessions to build models and analyse data.

The syllabus for this course is prepared by Avisek Gupta.

4. Design and Analysis of Algorithms

A. **Description**
The course introduces the basics of computational complexity analysis and various algorithm design paradigms. The goal is to provide students with solid foundations to deal with a wide variety of computational problems, and to provide a thorough knowledge of the most common algorithms and data structures. After the course, a student should be able to analyze the asymptotic performance of algorithms, write rigorous correctness proofs for algorithms, and apply important algorithmic design paradigms and methods of analysis.

B. **Pre-requisites**
Undergraduate level Mathematics, basic programming skills.

C. **Outline of the syllabus**

1. Worst and average case analysis. Recurrence relations. (03 hours)
2. Efficient algorithms for sorting, searching, and selection (06 hours)
3. Data structures: binary search trees, heaps, hash tables. (09 hours)
4. Algorithm design techniques: (12 hours) divide-and-conquer, dynamic programming, greedy algorithms, amortized analysis, randomization. Use problems from different domains including computational geometry.
5. Algorithms for fundamental graph problems (09 hours) minimum-cost spanning tree, connected components, topological sort, and shortest paths. (Possible additional topics: network flow, string searching.)
6. NP-complete problems. Approximation algorithms for NP-complete problems. (06 hours)

D. **Duration** 45 hours (15 weeks, 3 hours per week).

E. **Learning outcome and the objective of the course**
Students who complete the course will have the ability to do the following: (i) apply knowledge of computing and mathematics to algorithm design, (ii) analyze a problem and identify the computing requirements appropriate for its solution, (iii) design, implement, and evaluate an algorithm to meet desired needs, (iv) apply mathematical foundations, algorithmic principles, and computer science theory to model and design computer-based systems in a way that demonstrates comprehension of the trade-offs involved in design choices, (v) apply design and development principles in the construction of software systems of varying complexity, and (vi) use current techniques, skills, and tools necessary for computing practice.

F. **Books and References**

(a) T. H. Cormen, C. E. Leiserson and R. L. Rivest: Introduction to Algorithms, PrenticeHall of India, New Delhi, 1998.
(b) J. Kleinberg, E. Tardos: Algorithm Design, Pearson Education, 2006.
(c) A. Aho, J. Hopcroft and J. Ullman: The Design and Analysis of Computer Algorithms, A. W. L, International Student Edition, Singapore, 19983.
(d) S. Baase: Computer Algorithms: Introduction to Design and Analysis, 2nd ed., Addison-Wesley, California, 1988.

G. **Assessment methodology**
Written and Programming assignments, Examinations.

H. **Pedagogic methodology**
Lectures, presentations, and Programming sessions.

This course is proposed by Samiran Chattopadhyay.

5. Research Methodology

A. **Description:** The primary objective of this course is to enable the students, irrespective of their disciplines, in developing the most appropriate methodology for their research studies; and to make them familiar with the art of exploiting different research methods and techniques. The participants of the course should obtain a guideline on how to write, publish, present, and review scientific papers. The course aims to guide the students regarding the publication ethics and misconducts. It is expected that the course will assist in the accomplishment of exploratory as well as result-oriented research studies.

B. **Pre-requisites:** Knowledge of Mathematics at high school level.

C. **Outline of the syllabus:**

(a) **Fundamentals of Research**:

Basics: Definition, Purpose and Classification of research, Fundamentals of research methods, Writing a research proposal;

Problem Identification: Review of literature, Broadening knowledge base in the specific research area, Bringing clarity and focus to the research problem, Writing a research proposal, Writing research reports/papers;

Identifying variables: The difference between a concept and a variable, Converting concepts into variables, Types of variable, Types of measurement scale. (6× 1.5 hr=9 hr)

(b) **Research Design**:

Selecting a study design: Differences between quantitative and qualitative study designs, Study designs in quantitative research,

Data collection: Selecting a method of data collection, Differences in the methods of data collection in quantitative and qualitative research, Major approaches to information gathering, Methods of data collection in qualitative research. (4× 1.5 hr=6 hr)

(c) **Fundamentals of Statistics**: Frequency, Measure of Central Tendency, Dispersion, Regression and Interpretation of Results.

(10×1.5 hr=15 hr)

(d) **Basics of Probability**: Definition of Probability, Conditional Probability, Bayes' Random Variable, Probability Distribution. (6× 1.5 hr=9 hr)

(e) **Optimization Techniques**: Maxima & Minima, Condition of Optimality, Linear Programming Problem (Introduction, Formation of LPP, Graphical method of solution). (4×1.5 hr= 6 hr)

(f) **Research Ethics**:

Philosophy & Ethics: Introduction to Philosophy, definition, nature and scope, concepts, branches, nature of moral philosophy, nature of moral judgements and reactions, Ethics with respect to science and research, Intellectual honesty and research integrity;

Scientific misconducts: Falsification, Fabrication, and Plagiarism, Redundant publications; duplicate and overlapping publications, Selective reporting and misrepresentation of data, Conflict of interest;

Publication misconduct: parallel submission, authorship and contributorship, Ethical issues regarding the sponsoring organisation, Restrictions imposed by the sponsoring organisation, The misuse of information. (6× 1.5 hr= 9 hr)

(g) **Practical Session**: Application of software; Latex, Beamer, Presentation towards communication skill for professional development, Audience analysis and persuasion techniques, Use of plagiarism detection soft ware like Turnitin, Urkund and other open source soft ware tools. (4× 1.5 hr= 6 hr)

D. **Duration:** 60 hours (15 weeks, 4 hours per week).

E. **Learning outcome and the objective of the course:** Attending the course will reduce the probability for a student to be derailed from the research track. After completion of the course, a student should learn how to complete a project within a time bound in a scientific way. It is expected that the course can generate the awareness amongst the students about the publication ethics and publication misconducts. It is also expected that a student can write, present and communicate his/her research problem(s) professionally.

F. **References:**

(a) Research Methodology Methods & Techniques, C. R. Kothari, New Age International (P) Limited, Publisher, 2004.

(b) Research Methodology: A Step-by-Step Guide for Beginners (Fourth Edition), Ranjit Kumar, SAGE publishing, 2015.

(c) Fundamental of Research Methodology and Statistics, Y.K. Singh, New Age International (P) Limited, 2006.

(d) Research Methods and Statistics: A Critical Thinking Approach, Sherri L Jackson, American Psychological Association, 5th edition, Cengage Learning, 2014

(e) Research Ethics for Scientists: A Companion for Students, C. Neal Stewart Jr., Wiley Publishing, 2011.

(f) The student's guide to research ethics, Paul Oliver, Open University Press, McGraw-Hill Education, McGraw-Hill House, second edition, 2010.

G. **Assessment methodology:** Continuous assessment will be done through tutorials, assignments, quizzes, and group discussions. Weightage will be given for active participation. Written examination may also be conducted.

H. **Pedagogic methodology:** Class room teaching, guest lectures, group discussions, and practical sessions.

The syllabus for this course is prepared by Arpita Maitra.