

# IIT Madras BS Degree in Data Science and Applications

---

## Complete Syllabus

---

### Table of Contents

- [1. Foundational Level Courses](#)
  - [2. Diploma Level Courses](#)
  - [3. Projects](#)
  - [4. Degree / PGD / M.Tech Level Courses](#)
  - [5. Credit Transfer & Additional Information](#)
- 

### Foundational Level Courses

#### BSMA1001 — Mathematics for Data Science I

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Neelesh Upadhye (Dept. of Mathematics, IIT Madras), Prof. Madhavan Mukund (Director, Chennai Mathematical Institute), Prof. Sarang S Sane (Dept. of Mathematics, IIT Madras)

#### Course Description:

This course introduces functions (straight lines, quadratic, polynomials, exponentials and logarithms), calculus of one variable and discrete mathematics (basics, graph theory) with many examples. The students will be exposed to the idea of using abstract mathematical structures to represent concrete real life situations.

#### Syllabus:

- Week 1:** Set Theory – Number system, Sets and their operations, Relations and functions – Relations and their types, Functions and their types
- Week 2:** Rectangular coordinate system, Straight Lines – Slope of a line, Parallel and perpendicular lines, Representations of a Line, General equations of a line, Straight-line fit
- Week 3:** Quadratic Functions – Quadratic functions, Minima, maxima, vertex, and slope, Quadratic Equations
- Week 4:** Algebra of Polynomials – Addition, subtraction, multiplication, and division, Algorithms, Graphs of Polynomials – X-intercepts, multiplicities, end behavior, and turning points, Graphing & polynomial creation
- Week 5:** Functions – Horizontal and vertical line tests, Exponential functions, Composite functions, Inverse functions
- Week 6:** Logarithmic Functions – Properties, Graphs, Exponential equations, Logarithmic equations
- Week 7:** Sequence and Limits – Function of One variable – Function of one variable – Graphs and Tangents – Limits for sequences – Limits for function of one variable – Limits and Continuity
- Week 8:** Derivatives, Tangents and Critical points – Differentiability and the derivative – Computing derivatives and L'Hôpital's rule – Derivatives, tangents and linear approximation – Critical points: local maxima and minima
- Week 9:** Integral of a function of one variable – Computing areas, Computing areas under a curve, The integral of a function of one variable – Derivatives and integrals for functions of one variable
- Week 10:** Graph Theory – Representation of graphs, Breadth-first search, Depth-first search, Applications of BFS and DFS; Directed Acyclic Graphs – Complexity of BFS and DFS, Topological sorting
- Week 11:** Longest path, Transitive closure, Matrix multiplication Graph theory Algorithms – Single-source shortest paths, Dijkstra's algorithm, Bellman-Ford algorithm, All-pairs shortest paths, Floyd-Warshall algo-

rithm, Minimum cost spanning trees, Prim's algorithm, Kruskal's algorithm

- **Week 12:** Revision

---

## BSMA1002 — Statistics for Data Science I

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Usha Mohan (Dept. of Management Studies, IIT Madras)

### Course Description:

The students will be introduced to large datasets. Using this data, the students will be introduced to various insights one can glean from the data. Basic concepts of probability also will be introduced during the course leading to a discussion on Random variables.

### Syllabus:

- **Week 1:** Introduction and type of data, Descriptive and Inferential statistics, Scales of measurement
- **Week 2:** Describing categorical data, Frequency distribution of categorical data, Best practices for graphing categorical data, Mode and median for categorical variable
- **Week 3:** Describing numerical data – Frequency tables for numerical data, Measures of central tendency – Mean, median and mode, Quartiles and percentiles, Measures of dispersion – Range, variance, standard deviation and IQR, Five number summary
- **Week 4:** Association between two variables – Association between two categorical variables – Using relative frequencies in contingency tables, Association between two numerical variables – Scatterplot, Covariance, Pearson correlation coefficient, Point Bi-serial correlation coefficient
- **Week 5:** Basic principles of counting and factorial concepts – Addition rule of counting, Multiplication rule of counting, Factorials
- **Week 6:** Permutations and Combinations
- **Week 7:** Probability – Basic definitions of probability, Events, Properties of probability
- **Week 8:** Conditional probability – Multiplication rule, Independence, Law of total probability, Bayes' theorem
- **Week 9:** Random Variables – Random experiment, sample space and random variable, Discrete and continuous random variable, Probability mass function, Cumulative density function
- **Week 10:** Expectation and Variance – Expectation of a discrete random variable, Variance and standard deviation of a discrete random variable
- **Week 11:** Bernoulli trials, Independent and identically distributed Bernoulli trials, Binomial random variable, Expectation and Variance of a binomial random variable, Poisson distribution, Geometric distribution, Negative binomial distribution
- **Week 12:** Introduction to continuous random variables – Area under the curve, Properties of pdf, Uniform distribution, Exponential distribution, Normal distribution

---

## BSCS1001 — Computational Thinking

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Madhavan Mukund (Director, Chennai Mathematical Institute), Prof. G Venkatesh (Dept. of Humanities and Social Sciences, IIT Madras)

### Course Description:

The students will be introduced to a number of programming concepts using illustrative examples which will

be solved almost entirely manually. The manual execution of each solution allows for close inspection of the concepts being discussed.

### Syllabus:

- **Week 1:** Variables, Initialization, Iterators, Filtering, Datatypes, Flowcharts, Sanity of data
- **Week 2:** Iteration, Filtering, Selection, Pseudocode, Finding max and min, AND operator
- **Week 3:** Multiple iterations (non-nested), Three prizes problem, Procedures, Parameters, Side effects, OR operator
- **Week 4:** Nested iterations, Birthday paradox, Binning
- **Week 5:** List, Insertion sort
- **Week 6:** Table, Dictionary
- **Week 7:** Graph, Matrix
- **Week 8:** Adjacency matrix, Edge labelled graph
- **Week 9:** Backtracking, Tree, Depth First Search (DFS), Recursion
- **Week 10:** Object oriented programming, Class, Object, Encapsulation, Abstraction, Information hiding, Access specifiers
- **Week 11:** Message passing, Remote Procedure Call (RPC), Cache memory, Parallelism, Concurrency, Polling, Preemption, Multithreading, Producer Consumer, Atomicity, Consistency, Race condition, Deadlock, Broadcasting
- **Week 12:** Top-down approach, Bottom-up approach, Decision tree, Numerical prediction, Behaviour analysis, Classification

---

### BSHS1001 — English I

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Rajesh Kumar, Ms. Karthika Sathyanathan, Prof. Shreesh Chaudhary (Dept. of Humanities and Social Sciences, IIT Madras), Prof. Om Prakash (Dept. of Humanities and Social Sciences, Gautam Buddha University Noida)

### Course Description:

This course aims at achieving fluency and confidence in spoken and written English. This course will use insights from theories of learning and dominant methods of teaching language.

### Syllabus:

- **Week 1:** Sounds and Words (Vowel and Consonant sounds)
  - **Week 2:** Parts of Speech
  - **Week 3:** Sentences (Phrases and Idioms)
  - **Week 4:** Speaking Skills (Spoken English Preliminaries)
  - **Week 5:** Tenses and Agreement in English Sentences
  - **Week 6:** Reading Skills (Skimming, Scanning and Comprehension)
  - **Week 7:** Listening Skills
  - **Week 8:** Aspiration, Word Stress and Syllabification
  - **Week 9:** Speaking Skills (Presentation and Group Discussion)
  - **Week 10:** Grammar (Common Errors in English) and Writing Skills
  - **Week 11:** Writing Skills (Basics of Writing)
  - **Week 12:** Writing Skills (Professional Writing)
-

## BSMA1003 — Mathematics for Data Science II

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Sarang S Sane (Dept. of Mathematics, IIT Madras)

### Course Description:

This course aims to introduce the basic concepts of linear algebra, calculus and optimization with a focus towards the application area of machine learning and data science.

### Syllabus:

- **Week 1:** Vector and matrices – Vectors; Matrices; Systems of Linear Equations; Determinants (part 1); Determinants (part 2)
- **Week 2:** Solving linear equations – Determinants (part 3); Cramer's Rule; Solutions to a system of linear equations with an invertible coefficient matrix; The echelon form; Row reduction; The Gaussian elimination method
- **Week 3:** Introduction to vector spaces – Introduction to vector spaces; Some properties of vector spaces; Linear dependence; Linear independence – Part 1; Linear independence – Part 2
- **Week 4:** Basis and dimension – What is a basis for a vector space?; Finding bases for vector spaces; What is the rank/dimension for a vector space; Rank and dimension using Gaussian elimination
- **Week 5:** Rank and Nullity of a matrix; Introduction to Linear transformation – The null space of a matrix: finding nullity and a basis – Part 1; The null space of a matrix: finding nullity and a basis – Part 2; What is a linear mapping – Part 1; What is a linear mapping – Part 2; What is a linear transformation
- **Week 6:** Linear transformation, Kernel and Images – Linear transformations, ordered bases and matrices; Image and kernel of linear transformations; Examples of finding bases for the kernel and image of a linear transformation
- **Week 7:** Equivalent and Similar matrices; Introduction to inner products – Equivalence and similarity of matrices; Affine subspaces and affine mappings; Lengths and angles; Inner products and norms on a vector space
- **Week 8:** Orthogonality, Orthonormality; Gram-schmidt method – Orthogonality and linear independence; What is an orthonormal basis? Projections using inner products; The Gram-Schmidt process; Orthogonal transformations and rotations
- **Week 9:** Multivariable functions, Partial derivatives, Limit, continuity and directional derivatives – Multivariable functions: visualization; Partial derivatives; Directional derivatives; Limits for scalar-valued multivariable functions; Continuity for multivariable functions; Directional derivatives in terms of the gradient
- **Week 10:** Directional ascent and descent, Tangent (hyper) plane, Critical points – The directional of steepest ascent/descent; Tangents for scalar-valued multivariable functions; Finding the tangent hyper(plane); Critical points for multivariable functions
- **Week 11:** Higher order partial derivatives, Hessian Matrix and local extrema, Differentiability – Higher order partial derivatives and the Hessian matrix; The Hessian matrix and local extrema for  $f(x,y)$ ; The Hessian matrix and local extrema for  $f(x,y,z)$ ; Differentiability for Multivariable Functions; Review of Maths - 2

## BSMA1004 — Statistics for Data Science II

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Andrew Thangaraj (Electrical Engineering Department, IIT Madras)

### Course Description:

This second course will develop on the first course on statistics and further delve into the main statistical

problems and solution approaches.

**Syllabus:**

- **Week 1:** Multiple random variables – Two random variables, Multiple random variables and distributions
- **Week 2:** Multiple random variables – Independence, Functions of random variables – Visualization, functions of multiple random variables
- **Week 3:** Expectations – Casino math, Expected value of a random variable, Variance and standard deviation, Covariance and correlation, Inequalities
- **Week 4:** Continuous random variables, Discrete vs continuous, Weight data, Distribution functions, Density functions, Colab illustrations
- **Week 5:** Multiple continuous random variables – Height and weight data, Two continuous random variables, Averages of random variables – Colab illustration, Limit theorems, IPL data – histograms and approximate distributions, Jointly Gaussian random variables – Probability models for data – Simple models, Models based on other distributions, Models with multiple random variables, dependency, Models for IPL powerplay, Models from data
- **Week 6:** Refresher week
- **Week 7:** Estimation and Inference I
- **Week 8:** Estimation and Inference II
- **Week 9:** Bayesian estimation
- **Week 10:** Hypothesis testing I
- **Week 11:** Hypothesis Testing II
- **Week 12:** Revision week

---

**BSCS1002 — Programming in Python**

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Sudarshan Iyengar (Dept. of Computer Science and Engineering, IIT Ropar)

**Course Description:**

This will be the first formal programming course that students will see in this programme. The goal of this course is to introduce Python programming, which is used throughout the programme, with a basic problem solving and algorithmic flavour.

**Syllabus:**

- **Week 1:** Introduction to algorithms
  - **Week 2:** Conditionals
  - **Week 3:** Conditionals (Continued)
  - **Week 4:** Iterations and Ranges
  - **Week 5:** Iterations and Ranges (Continued)
  - **Week 6:** Basic Collections in Python
  - **Week 7:** Basic Collections in Python (Continued)
  - **Week 8:** Basic Collections in Python (Continued)
  - **Week 9:** File Operations
  - **Week 10:** File Operations (Continued)
  - **Week 11:** Module system in python
  - **Week 12:** Basic Pandas and Numpy processing of data
-

## BSHS1002 — English II

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Rajesh Kumar, Ms. Karthika Sathyanathan, Prof. Shreesh Chaudhary (Dept. of Humanities and Social Sciences, IIT Madras), Prof. Om Prakash (Dept. of Humanities and Social Sciences, Gautam Buddha University Noida)

### Course Description:

Focus on achieving greater degree of fluency in functional and conversational English to understand subtle and detailed meaning in conversations and texts through short literary pieces and contextualized content.

### Syllabus:

- **Week 1:** Patterns in Sentences
- **Week 2:** Patterns in Sentences (Continued)
- **Week 3:** Patterns in Sentences (Continued)
- **Week 4:** Listening Skills
- **Week 5:** Listening Skills (Continued)
- **Week 6:** Speaking Skills
- **Week 7:** Speaking Skills (Continued)
- **Week 8:** Reading Skills
- **Week 9:** Writing Skills
- **Week 10:** Writing Skills (Continued)
- **Week 11:** Social Skills
- **Week 12:** Social Skills (Continued)

## Diploma Level Courses

### BSCS2001 — Database Management Systems

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Partha Pratim Das (Dept. of Computer Science and Engineering, IIT Kharagpur)

### Course Description:

A comprehensive introduction to databases and PostgreSQL, database management, database design, and relevant topics like database security, integrity, and concurrency.

### Syllabus:

- **Week 1:** Course Overview
- **Week 2:** Relational Model and Basic SQL
- **Week 3:** Intermediate and Advanced SQL
- **Week 4:** Relational Query Languages and Database Design
- **Week 5:** Functional Dependency and Normal Forms
- **Week 6:** Functional Dependency and Normal Forms (cont.)
- **Week 7:** Application Development
- **Week 8:** Storage Management
- **Week 9:** Indexing and Hashing
- **Week 10:** Transactions
- **Week 11:** Backup and Recovery

- **Week 12:** Query Optimization and Conclusion

### BSCS2002 — Programming, Data Structures and Algorithms using Python

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Madhavan Mukund (Director, Chennai Mathematical Institute)

#### Course Description:

A good foundation course to introduce basic concepts in the design and analysis of algorithms as well as standard data structures, using Python as a base language for implementing these.

#### Syllabus:

- **Week 1:** Python Refresher
- **Week 2:** Complexity, Notations, Sorting and Searching Algorithms
- **Week 3:** Arrays, Lists, Stacks, Queues, Hashing
- **Week 4:** Graph Algorithms
- **Week 5:** Graph Algorithms (Continued)
- **Week 6:** Union-Find Data Structure, Priority Queue, Heap, Binary Search Tree
- **Week 7:** Balanced Search Tree, Greedy Algorithms
- **Week 8:** Divide and Conquer
- **Week 9:** Dynamic Programming
- **Week 10:** String or Pattern Matching Algorithms
- **Week 11:** Network Flows, Linear Programming, Class of Algorithms
- **Week 12:** Summary

### BSCS2003 — Modern Application Development I

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Nitin Chandrachoodan (Dept. of Electrical Engineering, IIT Madras), Thejesh GN (Independent Technologist, Multiple Affiliations)

#### Course Description:

Building a modern application involves many different aspects: front end, recording transactions, storage, connecting to a remote server, using APIs etc. The courses Modern Application Development I and II go through all these aspects through a detailed and evolving case study, teaching the relevant programming skills as the course progresses.

#### Syllabus:

- **Week 1:** Basic terminologies of Web
- **Week 2:** Web Pages written in HTML and CSS
- **Week 3:** Presentation layer – View
- **Week 4:** Models – Introduction to databases
- **Week 5:** Controllers – Business logic
- **Week 6:** APIs and REST APIs
- **Week 7:** Backend Systems
- **Week 8:** Application Frontend
- **Week 9:** Application Security

- **Week 10:** Testing of Web Applications
- **Week 11:** HTML Evolution and Beyond HTML
- **Week 12:** Application Deployment

### BSCS2005 — Programming Concepts using Java

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Madhavan Mukund (Director, Chennai Mathematical Institute)

#### Course Description:

This course uses Java to provide an understanding of core ideas in object oriented programming, exception handling, event driven programming, concurrent programming and functional programming.

#### Syllabus:

- **Week 1:** Basic Object Oriented Programming: Class Hierarchy
- **Week 2:** Basic Object Oriented Programming: Inheritance, Overriding
- **Week 3:** Basic Object Oriented Programming: Polymorphism
- **Week 4:** Basic Object Oriented Programming: Abstract Classes
- **Week 5:** Collections. Iterators.
- **Week 6:** Generics. Callbacks.
- **Week 7:** Cloning. I/O serializations. Packages
- **Week 8:** Cloning. I/O serializations. Packages (Continued)
- **Week 9:** Exception handling
- **Week 10:** Concurrent programming
- **Week 11:** Concurrent programming (Continued)
- **Week 12:** Concurrent programming (Continued)

### BSCS2006 — Modern Application Development II

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Nitin Chandrachoodan (Dept. of Electrical Engineering, IIT Madras), Thejesh GN (Independent Technologist, Multiple Affiliations)

#### Course Description:

Building a modern application involves many different aspects: front end, recording transactions, storage, connecting to a remote server, using APIs etc. The courses Modern Application Development I and II go through all these aspects through a detailed and evolving case study, teaching the relevant programming skills as the course progresses.

#### Syllabus:

- **Week 1:** Basics of JavaScript
- **Week 2:** Advanced JavaScript
- **Week 3:** Introduction to Web Frontend
- **Week 4:** Introduction to VueJS
- **Week 5:** Vue with APIs
- **Week 6:** Advanced Vuejs
- **Week 7:** Advanced State Management

- **Week 8:** Authentication and Designing APIs
- **Week 9:** Asynchronous Jobs
- **Week 10:** Inter-Service Messaging and Webhooks
- **Week 11:** Performance
- **Week 12:** Project

## BSSE2001 — System Commands

Attribute	Details
L-T-P-C	3-0-0-3
Instructor(s)	Prof. Gandham Phanikumar (Dept. of Metallurgical and Materials Engineering, IIT Madras)

### Course Description:

This course will take the student from the basics in command line usage to automating complex tasks – typically in a linux environment. The availability of WSL on Windows and the bash shell on MacOS ensure that the contents of this course are useful to every user with an ambition to get more done by their computer. Given a fairly complicated task at hand, one need not look for a perfect program that can achieve it. One can learn to use commands with their rich options, input / output redirections and combination with other commands to achieve the same task. Commands and declarations can be written with control structures in a script file to prepare shell scripts. These skills liberate the user from the need to pick any particular graphical user interface. A terminal is all that one needs to get most of the work done. There are also tasks that are difficult to attempt on a graphical interface while a shell script could execute that simpler and faster. By the end of the course, students will admire the elegance, brevity and efficiency of execution of tasks on the command line. Programmers can become more efficient and gain insights into the way their system works.

### Syllabus:

- **Week 1:** Introduction to GNU/Linux OS. Setting up and running Linux environment. The command line environment. Knowing hardware of your machine – Information commands such as `hwinfo`, `lshw`, `df`, `free` etc. Diagnostics – commands to fetch hardware information such as battery state, memory modules etc. Knowing the OS and software of your machine – Commands to get details about operating system, versions etc. Packages – installed / available. Input / output redirection.
- **Week 2:** Introduction to packages and repositories. Using 'apt' commands to manage packages. File types and related commands. Understanding file permissions and access modes. Managing file permissions through symbolic and numeric mode. Concept of environment variables. Important environment variables such as `HOME`, `USER` and `$PATH`
- **Week 3:** Managing shell variables. Prompt strings. Symbolic links and hard links, brief introduction to inode numbers. Exploring the root file system and related commands. Using shell shortcuts with commands. Slicing output. Managing programs currently running on the machine. Shell access to a local / remote machine.
- **Week 4:** Redirection to script, variable and for logging purpose. Using pipes. Introduction to regex; using regex patterns and `egrep`. Using `egrep` to extract useful information from files. `find` command and its uses – patterns to pick specific files in a folder, using `exec`. Command line editors (`nano`, `vi`, `emacs` – syntax highlighting & prompting, configuring options). Writing and running simple Bash scripts.
- **Week 5:** How are shell scripts interpreted? Using variables in scripts. Passing command line arguments to scripts, to create your own commands. More shell programming. Writing conditional statements using `if / else / fi`. Introduction to loops. Using functions. Configuring startup / periodic / recurring tasks
- **Week 6:** Text processing using AWK language. Using `awk` to run statistics on a data file. Using regex within `awk`. `Awk` as a programming language. Introduction to 'sed' – another text processing utility. Line by line processing to replace a regex pattern with a string. Use of place holders for matching regex patterns for use in replacing strings.

- **Week 7:** Introduction to make utility; concept of target and dependency; actions performed by make; conditional compilation; passing shell variables to make; File packaging utilities such as compress, tar, zip, gzip, bzip2, xz. Networking concepts – Introduction to IP addresses. Concept of localhost. Intranet and public addresses. Concept of ports and services that run on these ports. Concept of DNS and Domain names. Network diagnostics using tools and commands. Scripting a tool for analysis of logs.
- **Week 8:** Introduction to RAID for handling hardware failure. Introduction to version control. Git as a version control system: Overview of Git workflow, Branches, repositories, forks, etc. Creating and merging pull requests. Personal access tokens. Managing changes as local and remote. Working demo of a public repository using Git. Approval workflow. How a team collaborates on the private repository in an organization. Managing pull requests for the owner, raising issues and resolving them.

### BSCS2004 — Machine Learning Foundations

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Harish Guruprasad Ramaswamy, Prof. Arun Rajkumar, Prof. Prashanth LA (Dept. of Computer Sciences & Engineering, IIT Madras)

#### Course Description:

This course lays the groundwork for the upcoming ML courses by covering various fundamentals that do not necessarily fall under Machine Learning but are quite necessary for a comprehensive understanding of Machine Learning.

#### Syllabus:

- **Week 1:** Introduction to machine learning
- **Week 2:** Calculus
- **Week 3:** Linear Algebra – Least Squares Regression
- **Week 4:** Linear Algebra – Eigenvalues and eigenvectors
- **Week 5:** Linear Algebra – Symmetric matrices
- **Week 6:** Linear Algebra – Singular value decomposition, Principal Component Analysis in Image Processing
- **Week 7:** Unconstrained Optimisation
- **Week 8:** Convex sets, functions, and optimisation problems
- **Week 9:** Constrained Optimisation and Lagrange Multipliers. Logistic regression as an optimization problem
- **Week 10:** Examples of probabilistic models in machine learning problems
- **Week 11:** Exponential Family of distributions
- **Week 12:** Parameter estimation. Expectation Maximization

### BSMS2001 — Business Data Management

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. G Venkatesh, Prof. Suresh Babu (Dept. of Humanities and Social Sciences, IIT Madras), Dr. Milind Gandhe (Machine Intelligence and Robotics (MINRO) COE, IIIT Bangalore)

#### Course Description:

A significant source of data sets and problems for data scientists will come from the business domain. This course provides a basic understanding of how businesses are organised and run from a data perspective.

## Syllabus:

- **Week 1:** Consumption and demand: Micro & Macro economics: the role of data, production, consumption and exchange, consumption baskets, sources of consumer survey data
- **Week 2:** Micro-economic concepts: Utility: cardinal vs ordinal, indifference curves. Demand and supply curves, changes in demand and elasticity. production cost, cost curves. Make vs buy decisions, production quantity decisions
- **Week 3:** Firm level strategies and performance data: Objectives and types of pricing strategies, analysis of firm performance – key ratios. Analysis examples: Ultratech, Page Industries, Nestle, TCS
- **Week 4:** Analysing industry level data: Industry definition and classification codes, IIP and PMI, industry market structure and concentration indices, competitive positioning in an industry – Porter's five forces. Analysis examples: Cement industry, Textile industry, FMCG industry, IT industry
- **Week 5:** Case study 1 – Fabmart (E-Commerce): Introduction to E-Commerce, Fabmart case introduction, explanation of data set & questions to be answered, revenue pareto, volume pareto, scatter plot of sales and revenue, revenue trend
- **Week 6:** Fabmart case continued: Sales analysis, organisation of distribution centre, analysis of sales trends, average days of inventory, ledger, avoiding stockouts
- **Week 7:** Case study 2 – Ace Gears (Manufacturing): Introduction to the manufacturing sector, context of the automotive industry during the years 2019-2021, explanation of data set containing monthly information on sales, production, inventory and costing. Revenue trend analysis, portfolio management
- **Week 8:** Ace Gears case study continued: Regional sales analysis, sales agent planning, production scheduling, scrap analysis, unit level profitability analysis, raw material re-ordering and safety stock
- **Week 9:** Case study 3 – Tech Enterprises (IT): Introduction to HR as a function, Introduction to the Tech Enterprises, internal sourcing, ranking of internal candidates, job description, sourcing channels and their analysis, recruitment process and onboarding
- **Week 10:** Case study 4 – PayBuddy (Fin Tech): Introduction to Finance Industry and Fintech, payment processing and money flow, new credit product introduction, nudge economics, payment transaction and customer data set, identifying rules to target the appropriate customers
- **Week 11:** Paybuddy case continued: Introduction to A/B testing, analysis of the A/B testing data, credit risk evaluation, risk-return tradeoffs
- **Week 12:** Discussion on student acquired data sets. Wrap up (summary) of the case studies, course project work

---

## BSCS2007 — Machine Learning Techniques

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Arun Rajkumar (Dept. of Computer Science and Engineering, IIT Madras)

### Course Description:

To introduce the main methods and models used in machine learning problems of regression, classification and clustering. To study the properties of these models and methods and learn about their suitability for different problems.

### Syllabus:

- **Week 1:** Introduction; Unsupervised Learning – Representation learning – PCA
- **Week 2:** Unsupervised Learning – Representation learning – Kernel PCA
- **Week 3:** Unsupervised Learning – Clustering – K-means/Kernel K-means
- **Week 4:** Unsupervised Learning – Estimation – Recap of MLE + Bayesian estimation, Gaussian Mixture Model – EM algorithm
- **Week 5:** Supervised Learning – Regression – Least Squares; Bayesian view

- **Week 6:** Supervised Learning – Regression – Ridge/LASSO
- **Week 7:** Supervised Learning – Classification – K-NN, Decision tree
- **Week 8:** Supervised Learning – Classification – Generative Models – Naive Bayes
- **Week 9:** Discriminative Models – Perceptron; Logistic Regression
- **Week 10:** Support Vector Machines
- **Week 11:** Ensemble methods – Bagging and Boosting (Adaboost)
- **Week 12:** Artificial Neural networks: Multiclass classification

### BSCS2008 — Machine Learning Practice

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Ashish Tendulkar (Google AI, Google)

#### Course Description:

This companion course to the ML Theory course introduces the student to scikit-learn, a popular Python machine learning module, to provide hands-on problem solving experience for all the methods and models learnt in the Theory course.

#### Syllabus:

- **Week 1:** End-to-end machine learning project on scikit-learn
- **Week 2:** Graph Theory (VOL 3)
- **Week 3:** Regression on scikit-learn – Linear regression, Gradient descent – batch and stochastic
- **Week 4:** Polynomial regression, Regularized models
- **Week 5:** Logistic regression
- **Week 6:** Classification on scikit-learn – Binary classifier
- **Week 7:** Classification on scikit-learn – Multiclass classifier
- **Week 8:** Support Vector Machines using scikit-learn
- **Week 9:** Decision Trees, Ensemble Learning and Random Forests
- **Week 10:** Decision Trees, Ensemble Learning and Random Forests (Continued)
- **Week 11:** Neural networks models in scikit-learn
- **Week 12:** Unsupervised learning

### BSMS2002 — Business Analytics

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Rahul R Marathe (Dept. of Management Studies, IIT Madras)

#### Course Description:

The problems faced by decision makers in today's business environments are extremely complex. Hence, the task of making good decisions is not easy. The answer is in building quantitative models, and this course is designed to help you understand the fundamentals of this critical, foundational, business skill. The business application of statistical methods is the core focus of this course. In that sense, the course builds on the core course in the first year of the program. That basic course focused on the preliminaries of the area. This course highlights a business application and then demonstrates an application of a statistical technique to solve that scenario and arrive at the best decisions and insights.

#### Syllabus:

- **Week 1:** Data dashboarding – Insights from data summary
- **Week 2:** Can summarizing the data provide insights?
- **Week 3:** Do people in different cities prefer different brands?
- **Week 4:** Predicting the stock returns – Regression basics
- **Week 5:** How do you pay a professor? – Regression diagnostics – Path variables
- **Week 6:** Can I cure cancer? – Logistic Regression – Connection with classification problem
- **Week 7:** What is the impact of repeatedly watching the same ad? – Repeated measures ANOVA
- **Week 8:** When the data has a time axis: Time series modeling

### BSSE2002 — Tools in Data Science

Attribute	Details
L-T-P-C	3-0-0-3
Instructor(s)	Prof. S Anand (CEO, Gramener)

#### Course Description:

This course will teach students to use popular tools for sourcing data, transforming it, building and optimizing models, communicating these as visual stories, and deploying them in production.

#### Syllabus:

- **Week 1:** Discover the data
- **Week 2:** Get the data
- **Week 3:** Prepare the data
- **Week 4:** Model the data
- **Week 5:** Modern tools to simplify deep learning models
- **Week 6:** Design your output
- **Week 7:** Narrate a story
- **Week 8:** Deploy the results

### BSDA2001 — Introduction to Deep Learning and Generative AI

Attribute	Details
L-T-P-C	4-0-0-4
Co-requisite	CS2008: Machine Learning Practice
Instructor(s)	Prof. Balaji Srinivasan (Professor, Dept. of Mechanical Engineering, Wadhvani School of AI, IIT Madras), Prof. Ganapathy (Professor, Wadhvani School of AI, IIT Madras)

#### Course Description:

This course aims to provide a comprehensive introduction to the foundational and practical aspects of Deep Learning and Generative AI. Through a balanced blend of theoretical concepts and hands-on experience, students will learn to build, train, and evaluate artificial neural networks for a variety of tasks in computer vision and natural language processing. The course covers key architectures such as Convolutional Neural Networks (CNNs) for image data, Recurrent Neural Networks (RNNs) and LSTMs for sequential data, and extends into the realm of generative models including Autoencoders, Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs) and Large Language Models (LLMs). By the end of the course, learners will gain the skills to implement core deep learning models and apply generative AI techniques to solve practical problems.

#### Syllabus:

- **Week 1: Artificial Neural Networks – Theory**  
Introduction to Deep Learning, Artificial neurons, neural networks, layers, Activation functions and loss metrics
- **Week 2: Artificial Neural Networks – Practice**  
Hands-on: Build simple neural networks using TensorFlow/Keras, Experimentation with activation functions and optimization methods
- **Week 3: Modeling Vision — CNN – Theory**  
Introduction to Convolutional Neural Networks (CNNs), CNN architecture basics, convolution and pooling layers
- **Week 4: Modeling Vision — CNN – Practice**  
Hands-on: CNN-based image classification (e.g., MNIST, CIFAR-10)
- **Week 5: Modeling Sequential Data – Theory**  
Sequence models, Recurrent Neural Networks (RNNs), LSTMs
- **Week 6: Modeling Sequential Data – Practice**  
Hands-on: Sentiment analysis with LSTM, Text generation with simple RNN/LSTM models, Time series prediction tasks
- **Week 7: Generative AI for vision — Variational AutoEncoders and GANs – Theory**  
Introduction to Generative AI, Autoencoders and Variational Autoencoders (VAEs) basics, GANs
- **Week 8: Generative AI for vision — Diffusion Models**  
Diffusion Probabilistic models, training and inference
- **Week 9: Generative AI for vision — Practice**  
GANs and Pretrained Diffusion Models to generate Images
- **Week 10: Large Language Models — Transformer Architecture**  
Word Embeddings, Tokenization, Attention
- **Week 11: Large Language Models – encoder, decoder, encoder decoder models**  
BERT like models for NLP tasks, Decoders for Text Generation, Machine Translation, Fine-Tuning LLMs (PEFT, LoRA)
- **Week 12: Large Language Models – Practice**  
Prompting Techniques, Prompt Fine-tuning and other methods

## Projects

Project Code	Project Title	Credits
BSMS2001P	Business Data Management Project	2
BSCS2008P	Machine Learning Practice Project	2
BSCS2003P	Application Development 1 Project	2
BSCS2006P	Application Development 2 Project	2
BSDA2001P	Introduction to DL and GenAI Project	2

## Degree / PGD / M.Tech Level Courses

### BSCS3001 — Software Engineering

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Sridhar Iyer (Interdisciplinary Program in Educational Technology, IIT Bombay), Dr. Prajish Prasad (Dept. of Computer Science, FLAME University)

**Course Description:**

To prepare students to develop the essential skills required to become effective software engineers by introducing them to fundamental concepts in developing software, and essential practices employed by software developers, such as requirement gathering, creating software conceptual designs, software comprehension, debugging, testing and deployment.

**Syllabus:**

- **Week 1:** Deconstructing the software development process
- **Week 2:** Identify different types of software requirements (functional, non-functional)
- **Week 3:** Software Conceptual Design
- **Week 4:** Software Usability
- **Week 5:** Software Design – Modeling and Architecture
- **Week 6:** Software Design – Quality and Evaluation
- **Week 7:** Software Development – Program Comprehension
- **Week 8:** Software Development – Program Debugging
- **Week 9:** Software – Code Reviewing and Documentation
- **Week 10:** Software Testing
- **Week 11:** Software Deployment and Monitoring
- **Week 12:** Conclusion and other Aspects: Communication, Productivity and Organizations

**BSCS3002 — Software Testing**

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Meenakshi D'Souza (Dept. of Computer Science and Engineering, IIT Bangalore)

**Course Description:**

To prepare the students to understand the phases of testing based on requirements for a project, to apply the concepts taught in the course to formulate test requirements precisely, to design and execute test cases as a part of a standard software development IDE, and to apply specially designed test case design techniques for specific application domains.

**Syllabus:**

- **Week 1:** Introduction, testing in software development life-cycle, software testing process levels, testing terminologies, test automation
- **Week 2:** Techniques and algorithms for test case design:
  - Graphs based testing: Structural coverage criteria, data flow coverage criteria, Graph coverage for source code, design elements and specifications
  - Logic based testing: Predicates and clauses, coverage criteria based on logic expressions, Specification-based logic coverage, Logic coverage for finite state machines, Symbolic testing and concolic testing
  - Input space partitioning: Input domain modeling, combination strategies criteria, decision tables
  - Syntax based testing: Coverage criteria based on syntax, mutation testing
- **Week 3:** Open-source tools and frameworks for testing
- **Week 4:** Integration testing
- **Week 5:** Testing of applications: Testing OO-applications, web applications
- **Week 6:** Agile testing, Test driven development
- **Week 7:** Overview of non-functional testing techniques
- **Week 8:** Regression testing

---

## BSCS3003 — AI: Search Methods for Problem Solving

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Deepak Khemani (Dept. of Computer Science and Engineering, IIT Madras)

### Course Description:

To get a brief historical and philosophical perspective on artificial intelligence. To formulate problems in a general problem-solving framework. Obtain knowledge of domain independent search-based problem-solving algorithms. Obtain knowledge of stochastic, local, and population-based search algorithms. Lay the foundations of problem decomposition and rule-based methods. To implement game playing algorithms. To understand the relation between search methods and other formulations including planning, constraints and logical reasoning.

### Syllabus:

- **Week 1:** Introduction and philosophy. The Turing Test. The Winograd Schema Challenge. Placing search in the landscape of AI
- **Week 2:** Search spaces. Examples. State space search. Depth First, Breadth First, Iterative Deepening. Analysis
- **Week 3:** Heuristic search. Heuristic functions. Solution space search. Escaping local optima. Stochastic local search
- **Week 4:** Population based methods. Genetic Algorithms, emergent systems, Ant Colony Optimization
- **Week 5:** Finding optimal paths. Algorithm A\*. Admissibility of A\*
- **Week 6:** The monotone condition. Space saving versions of A\*. Sequence alignment
- **Week 7:** Game playing. Board games. Algorithms Minimax, Alpha-Beta, and SSS\*
- **Week 8:** Automated domain independent planning. Goal Stack Planning, Partial Order Planning
- **Week 9:** Problem decomposition with goal trees. Algorithm AO\*
- **Week 10:** Pattern directed inference systems. Forward chaining inference engine. The Rete algorithm
- **Week 11:** Constraint processing. Algorithm Backtracking. Arc consistency. Combining search and reasoning. Waltz algorithm. Model based diagnosis
- **Week 12:** Revision / Applications

---

## BSCS3004 — Deep Learning

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Mitesh Khapra (Dept. of Computer Science and Engineering, IIT Madras)

### Course Description:

To study the basics of Neural Networks and their various variants such as the Convolutional Neural Networks and Recurrent Neural Networks, to study the different ways in which they can be used to solve problems in various domains such as Computer Vision, Speech and NLP.

### Syllabus:

- **Week 1:** History of Deep Learning, McCulloch Pitts Neuron, Thresholding Logic, Perceptron Learning Algorithm and Convergence
- **Week 2:** Multilayer Perceptrons (MLPs), Representation Power of MLPs, Sigmoid Neurons, Gradient Descent
- **Week 3:** Feedforward Neural Networks, Representation Power of Feedforward Neural Networks, Backpropagation

- **Week 4:** Gradient Descent (GD), Momentum Based GD, Nesterov Accelerated GD, Stochastic GD, Adagrad, AdaDelta, RMSProp, Adam, AdaMax, NAdam, learning rate schedulers
- **Week 5:** Autoencoders and relation to PCA, Regularization in autoencoders, Denoising autoencoders, Sparse autoencoders, Contractive autoencoders
- **Week 6:** Bias Variance Tradeoff, L2 regularization, Early stopping, Dataset augmentation, Parameter sharing and tying, Injecting noise at input, Ensemble methods, Dropout
- **Week 7:** Greedy Layer Wise Pre-training, Better activation functions, Better weight initialization methods, Batch Normalization
- **Week 8:** Learning Vectorial Representations Of Words, Convolutional Neural Networks, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet
- **Week 9:** Visualizing Convolutional Neural Networks, Guided Backpropagation, Deep Dream, Deep Art, Fooling Convolutional Neural Networks
- **Week 10:** Recurrent Neural Networks, Backpropagation Through Time (BPTT), Vanishing and Exploding Gradients, Truncated BPTT
- **Week 11:** Gated Recurrent Units (GRUs), Long Short Term Memory (LSTM) Cells, Solving the vanishing gradient problem with LSTM
- **Week 12:** Encoder Decoder Models, Attention Mechanism, Attention over images, Hierarchical Attention, Transformers

### BSGN3001 — Strategies for Professional Growth

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Sivakumar M S (Dept. of Applied Mechanics, IIT Madras), Prasanna G, Kartic Vaidyanathan, Lt Col Jayakumar, Shiva Subramaniam (Guest Faculty, IIT Madras), Prof. Prathap Haridoss (Dept. of Metallurgical and Materials Engineering, IIT Madras), Senthil A V (Certified Financial Planner)

#### Course Description:

The course titled, 'Strategies for Professional Growth' prepares students for corporate life and to be global citizens and provides an opportunity to develop an all-around personality that would help future personal and professional growth. To enable the student to use the creative process to identify and solve problems in an effective way, to use structured creative thinking tools to investigate a particular matter from a variety of perspectives with clarity, to communicate and share thoughts/information accurately and effectively to understand each other, to become a team player, to value other cultures, to overcome obstacles she/he may face when performing a task, to work with hands to become better at engaging and enhancing their thought process, to get pertinent information crucial for learning about something or for communicating, to understand how decisions at the individual level or at the business level affect optimal utilisation of resources, to use conflict resolution tools to effectively resolve conflicts, to perceive emotions of themselves and of others and manage for a better outcome under various circumstances.

#### Syllabus:

- **Week 1:** Teamwork and Getting Along
- **Week 2:** Communication and Listening Skills
- **Week 3:** Cultivating a Growth Mindset
- **Week 4:** Leadership Lessons
- **Week 5:** Emotional Intelligence and Conflict Management
- **Week 6:** Systems Thinking
- **Week 7:** Engineering Sense
- **Week 8:** Fiscal and Economic Sense
- **Week 9:** Cross-cultural Understanding and Personal Grooming

- **Week 10:** Creativity and Thinking Skills
- **Week 11:** Presentation Skills
- **Week 12:** Summary

### BSBT4001 — Algorithmic Thinking in Bioinformatics

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Manikandan Narayanan (Dept. of Computer Science & Engineering, IIT Madras)

#### Course Description:

To prepare students to develop an algorithmic thinking to address key data science challenges in bioinformatics, to acquire knowledge of various problem formulations and algorithm paradigms, which have transformed the field of biomedicine in modern times, to obtain insights into many key bioinformatics algorithms on strings, trees, and graphs, many of which can be applied to other areas as well.

#### Syllabus:

- **Week 1:** Why computational biology?
- **Week 2:** Where in the Genome Does DNA Replication Begin? – Algorithmic warmup (frequent exact/inexact k-mers in a string)
- **Week 3:** Which DNA Patterns Play the Role of Molecular Clocks? – Randomized Algorithms (randomized motif search, Gibbs sampling)
- **Week 4:** How Do We Assemble Genomes? – Graph Algorithms (Eulerian paths, de Bruijn graphs)
- **Week 5:** How Do We Compare Biological Sequences? – Dynamic Programming (edit distance, single/multiple sequence alignment)
- **Week 6:** Which Animal Gave Us SARS? – Evolutionary Tree Reconstruction (distance-based phylogeny, neighbor-joining algorithm)
- **Week 7:** How Did Yeast Become a Winemaker? – Clustering Algorithms (hard and soft k-means)
- **Week 8:** How Do We Locate Disease-Causing Mutations? – Combinatorial Pattern Matching (suffix trees/arrays, Burrows-Wheeler transform)
- **Week 9:** Why Have Biologists Still Not Developed an HIV Vaccine? – Hidden Markov Models (Viterbi and forward-backward algorithms)
- **Week 10:** Was T. rex Just a Big Chicken? – Computational Proteomics (peptide identification and spectral match)
- **Week 11:** Which Motifs Are Hidden in a Biological Network? – Randomized Algorithms (colour coding for long paths in graphs)

### BSBT4002 — Big Data and Biological Networks

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Nirav Bhatt, Prof. Karthik Raman, Prof. Himanshu Sinha (Dept. of Biotechnology, IIT Madras)

#### Course Description:

To enable the students to "understand" biological data, to represent, and analyze various datasets from a network perspective, to encourage network thinking applied to problems across disciplines, to understand various network models used to model real-world networks, to apply network analytics techniques to understand biological networks, to implement basic network analysis algorithms in Python, to learn different AI/ML problem formulations for biological data, and to apply AI/ML techniques for analysis of biological data using Python.

**Syllabus:**

- **Week 1:** Introduction to Biological Big Data. Information Flow in Biological Systems
- **Week 2:** Omics datasets: Various flavours of big biological datasets (genomic, transcriptomic, proteomic, metabolomic, etc.)
- **Week 3:** Introduction to Graph theory. History. Types of graphs. Representing biological networks
- **Week 4:** Network structure: Key parameters, measures of centrality
- **Week 5:** Key Network Models: Erdos-Renyi, Watts-Strogatz (small-world) and Barabasi-Albert (power-law models)
- **Week 6:** Network clustering/community detection. Identifying motifs in networks. Studying network perturbations
- **Week 7:** Applications of network biology: Predicting drug targets, predicting drug molecules, synthesis of new molecules (chemoinformatics)
- **Week 8:** Applications of network biology: Epidemiology, Centrality-lethality hypothesis
- **Week 9:** AI & ML for Biological Data Analysis. Introduction to AI & ML tasks in biological networks
- **Week 10:** Biological network reconstruction from omics and literature data
- **Week 11:** Property prediction using network data. Node classification and link prediction
- **Week 12:** Analysis of heterogeneous and multi-layer/multiplex networks. Future Perspectives

---

**BSCS4001 — Data Visualization Design**

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Venkat (Industrial Design Center, IIT Bombay)

**Course Description:**

To provide students with the foundations necessary for understanding and extending the current state of the art in data visualization, to gain an understanding of the key techniques and theory used in visualization, including data models, graphical perception and techniques for visual encoding and interaction, to plan for data-based storytelling through charts, maps, and diagrams, to use visualization tools to transform quantitative information to visual representation, and to gain practical experience building and evaluating visualizations.

**Syllabus:**

- **Week 1:** Information visualization overview, historical perspective
  - **Week 2:** Vision, perception and cognition principles of information visualization
  - **Week 3:** Data principles and models
  - **Week 4:** Analysis and Insights
  - **Week 5:** Geo-visualization
  - **Week 6:** Map Abstraction
  - **Week 7:** Visual encoding of data
  - **Week 8:** Visualization design
  - **Week 9:** Data Stories
  - **Week 10:** Project presentations, Wrap up and course summary
-

## BSEE4001 — Speech Technology

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Umesh (Dept. of Electrical Engineering, IIT Madras), Prof. Hema A Murthy (Dept. of Computer Science and Engineering, IIT Madras)

### Course Description:

To understand the concepts of speech and speech technologies, and to apply them to real-world scenarios. To gain hands-on experience of the relevant toolkits used for speech processing.

### Syllabus:

- **Week 1:** Review of Signals and Systems, Continuous time signals and transforms, Discrete time signals, Discrete Fourier transform, Autocorrelation and Cross-Correlation
- **Week 2:** Acoustic Feature Analysis of Speech Signals I, II; Gaussian mixture models (GMM), universal background model (UBM-GMM), singular value decomposition (SVD)
- **Week 3:** Hidden Markov model (HMM), Examples of HMM based approach for ASR, TTS, speaker diarization; Information bottleneck (IB) based clustering for diarization
- **Week 4:** Introduction and History of ASR and TTS; Components of ASR: Acoustic Modelling, Punctuation Model (Lexicon) and language modelling (N-Gram Language models)
- **Week 5:** HMMs for Acoustic Modelling – Monophone, Triphone; Speech Synthesis: unit selection, statistical parametric synthesis (HTS)
- **Week 6:** Neural networks for building speech technologies; NN for Acoustic Modelling – Hybrid modelling – Hybrid-NN: DNN, CNN, TDNN
- **Week 7:** End-to-End Approaches I: CTC, Encoder-decoder Architecture E2E with RNN
- **Week 8:** Applications to ASR and TTS; End-to-End Approaches II
- **Week 9:** Encoder-decoder Architecture E2E with transformers for ASR and TTS; Interesting Problems
- **Week 10:** Speaker recognition/verification: with ivector, xvector; Speaker diarization: using x-vector
- **Week 11:** Speaker adaptation: (revisit i, x vectors) and introduce s-vectors; Code Switched Speech recognition; Speech Translation
- **Week 12:** Singing voice synthesis; voice conversion; generic voice synthesis

## BSMS4002 — Design Thinking for Data-Driven App Development

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Bala Ramadurai (Consultant and Guest Faculty, IIT Madras), Prof. Prathap Haridoss (Professor, Dept. of Metallurgical and Materials Engineering, IIT Madras)

### Course Description:

To enable the students to apply agile method to developing software, to design an app using the principles of design thinking, to develop an app for android, to collaborate with other developers using git version control method, and to learn the basics of marketing and customer support through their website.

### Syllabus:

- **Week 1:** Introduction to Design Thinking – Course Outline and Projects, Intro to Design from Everyday Life, Intro to Design Thinking in Software Apps, Project Management
- **Week 2:** Empathize – P1 – Persona Creation, Customer Journey Mapping, Demo on a real problem
- **Week 3:** Empathize – P1 – List of problems, How Might We questions, Demo on a real problem
- **Week 4:** Analyze – P1 – Multi-Why, Conflict of Interest, Demo on a real problem

- **Week 5:** Solve – P1 – Silent brainstorming, Inventive Principles, Concept creation, Demo on a real problem
- **Week 6:** Test – P1, Empathize – P2 – Assumptions, Features, Field trials
- **Week 7:** Website Development, Basics of Digital Marketing, User Experience Design
- **Week 8:** Analyze – P2
- **Week 9:** Solve – P2
- **Week 10:** Test – P2, Empathize – P3 – Obtaining insights/feedback from the customers or target users
- **Week 11:** Analyze – P3
- **Week 12:** Test – P3, Launch of the App

### BSMS4001 — Industry 4.0

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Rahul Marathe, Dr. Amit (Dept. of Management Studies, IIT Madras)

#### Course Description:

To describe various facets of Industry 4.0, to connect questions raised by Industry 4.0 with appropriate data science techniques, to develop data science tools for Industry 4.0, and to build data-centric business models.

#### Syllabus:

- **Week 1:** Introduction to Industry 4.0 – Evolution and history
- **Week 2:** Pillars of Industry 4.0
- **Week 3:** Industry 4.0 – India context
- **Week 4:** Supplier selection as a classification problem
- **Week 5:** Manufacturing 4.0
- **Week 6:** Prognosis
- **Week 7:** Quality 4.0
- **Week 8:** Inventory Optimization
- **Week 9:** Dynamic Pricing
- **Week 10:** Logistics 4.0
- **Week 11:** Future of Manufacturing – Business Focus on new paradigm
- **Week 12:** Next decade of Industry 4.0

### BSMS4003 — Financial Forensics

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Arun Kumar G (Dept. of Management Studies, IIT Madras)

#### Course Description:

An introduction to Finance and Accounting, The life cycle of a financial transaction, Areas where AI/ML is used in the Finance Industry, Importance of Model Explainability in the Regulated World, An introduction to solving real world finance problems – Credit Card Fraud Detection, Identity Fraud Detection, Anti Money Laundering Scenarios.

#### Syllabus:

- **Week 1:** INTRODUCTION TO ACCOUNTING AND FINANCE: Financial Statements, Uses relevance of financial statements, Decision making in the financial arena, Need for finance and accounting in a business world

- **Week 2:** FRAUD VULNERABILITIES AND FINANCE: Types of fraud, Why do they occur, Where do they occur, Detecting red flags
- **Week 3:** FORENSIC ACCOUNTING – 1: Intro to Forensic Accounting, Source of assignments and referrals for a forensic accountant, Role of the forensic accountant as an expert – identify any conflicts
- **Week 4:** FORENSIC ACCOUNTING – 2: Scope of forensic accounting, Process of forensic accounting – Analysis
- **Week 5:** BENFORD LAW AND IMPLEMENTATION: Benford Law, Using Benford Law in Excel to detect audit fraud
- **Week 6:** AGING ANALYSIS, PARETO AND OUTLIERS: Detecting fraud using Aging Analysis in Excel, Creating a Pareto Chart, Outlier identification
- **Week 7:** FINANCIAL TRANSACTION LIFECYCLE: Transaction Lifecycle, Where is AI/ML used in the present world?
- **Week 8:** ENTITY RESOLUTION: What is an entity in a transaction?, Entity resolution
- **Week 9:** ANOMALY DETECTION – 1: Supervised Anomaly Detection
- **Week 10:** ANOMALY DETECTION – 2: Unsupervised Anomaly Detection
- **Week 11:** Model Explainability: Financial Data Visualization using Tableau

### BSMS3002 — Market Research

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Ashok Sankethi (Kaybase Research)

#### Course Description:

To provide a basic understanding of research methodology and its implementation in different business domains, to understand the role, scope, process, cost, and value of marketing research, to match research techniques to marketing problems, to analyse data and translate them into actionable findings, to enable students to do hands-on research to solve business problems.

#### Syllabus:

- **Week 1:** The Role of Research in Marketing and the Marketing Research Process, Defining the Marketing Research Problem and Developing an Approach
- **Week 2:** Research Design, Exploratory Research Design, Descriptive Research Design, Causal Research Design and Test Markets
- **Week 3:** Surveys and Interviews, Measurement, Measurement Scales, Questionnaires, and Instruments
- **Week 4:** Key marketing issues like new product development, STP, branding, etc. and the MR tools and techniques to address each
- **Week 5:** Entering the Data in SPSS, Examining the Data and Univariate Data Analysis: Descriptive Statistics, Cross Tabulation, Graphical Display of Data, Chart Deception and Hypothesis Testing: Mean Differences, ANOVA, Multivariate Data Analysis: An Overview, Exploratory Factor Analysis, Cluster Analysis, Multiple Regression in SPSS

### BSDA5001 — Introduction to Big Data

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Rangarajan Vasudevan (Co-Founder & Chief Data Officer, Lentre.ai)

**Course Description:**

This course will introduce students to practical aspects of analytics at a large scale, i.e. big data. The course will start with a basic introduction to big data and cloud concepts spanning hardware, systems and software, and then delve into the details of algorithm design and execution at large scale.

**Syllabus:**

- **Week 1:** Introduction: Big data concepts & GCP Platform Setup
- **Week 2:** Cloud concepts: Cloud-Native architecture, serverless computing, message queues, PaaS, SaaS, IaaS
- **Week 3:** Types of Data: Data formats, sources & their semantics, processing & storage options on Cloud. Use of serverless to get started (e.g. Google Cloud Functions)
- **Week 4:** Intro to Big Data Engineering: Hadoop and PySpark
- **Week 5:** ETL: ETL, processing patterns for large data, ETL vs ELT, role of a scheduler
- **Week 6:** SQL & NoSQL: For most analysis tasks, SQL is sufficient. Tools like Spark SQL allow that familiarity to translate to big data solutions. Types of NoSQL, evolution, best-of-fit options
- **Week 7:** Streaming: Overview, Fundamental Concepts, Walkthrough of Google Pub/Sub & Google DataFlow as example technologies
- **Week 8:** Streaming: Kafka as another example of message queue technology & Spark Streaming
- **Week 9:** Big Data ML: DataProc with ML – including Spark ML (Batch processing)
- **Week 10:** Deep Learning with big data on cloud
- **Week 11:** Prep week for final project, summarizing key concepts, and also for Q&A and clarifications

**BSCS4003 — Privacy and Security in Online Social Media**

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Ponnurangam Kumaraguru (Dept. of Computer Science, IIIT Hyderabad)

**Course Description:**

Lists Various Privacy and security concerns on Online social Networks and described different methodologies used for solving security and privacy problems on Online social networks.

**Syllabus:**

- **Week 1:** Topics – Assessment, Social network analysis 101
- **Week 2:** Collecting data from social media – Collect data, analyze, and report
- **Week 3:** Text analysis of social media data – Analyze data and report
- **Week 4:** Cyber crime on social media – Analyze data and report
- **Week 5:** Cyber crime on social media – Reading paper and report
- **Week 6:** Fake news on social media – Analyze data and report
- **Week 7:** Fake news on social media – Reading paper and report
- **Week 8:** Privacy on social media – Analyze data and report
- **Week 9:** Privacy on social media – Reading paper and report
- **Week 10:** Ethics, bias on online social media – Reading paper and report
- **Week 11:** Computational social science – Reasoning / Comprehension of ideas
- **Week 12:** Computational social science on online social media – Reading paper and report

## BSMA2001 — Mathematical Thinking

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Amritanshu Prasad, Prof. Sankaran Viswanath (The Institute of Mathematical Science)

### Course Description:

To introduce ideas of proofs and problem solving in mathematics and to help in the transition from learning basic mathematical methods to the learning of more advanced mathematical methods and ideas.

### Syllabus:

- **Week 1:** Numbers, Sets and Functions
- **Week 2:** Language and Proofs
- **Week 3:** Induction, Bijections and Cardinality
- **Week 4:** Combinatorial Reasoning
- **Week 5:** Divisibility, Modular Arithmetic
- **Week 6:** The Rational Numbers
- **Week 7:** Two Principles of Counting
- **Week 8:** Recurrence Relations

---

## BSMA3012 — Linear Statistical Models

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Siva Athreya (International Centre for Theoretical Sciences – TIFR and Indian Statistical Institute, Bangalore Centre)

### Course Description:

To introduce linear statistical models and their applications in estimation and testing. The course will illustrate concepts with specific examples, data sets and numerical exercises using statistical package R.

### Syllabus:

- **Week 1:** Review of Estimation, Hypothesis Testing
  - **Week 2:** Review of working with R-package
  - **Week 3:** Least square estimation, estimable linear functions
  - **Week 4:** Normal equations
  - **Week 5:** Best Linear Unbiased Estimates (BLUEs)
  - **Week 6:** Gauss-Markov Theorem
  - **Week 7:** Degrees of freedom. Fundamental Theorems of Least Square
  - **Week 8:** Testing of linear hypotheses
  - **Week 9:** One-way and two-way classification models
  - **Week 10:** ANOVA and ANCOVA
  - **Week 11:** Q&A and clarifications
  - **Week 12:** Introduction to random effect models
-

## BSMA3014 — Statistical Computing

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dootika Vats (Assistant Professor, Dept. of Mathematics and Statistics, IIT Kanpur)

### Course Description:

To introduce computational methods involved in statistical estimation and learning problems.

### Syllabus:

- **Week 1:** Introduction to R, Introduction to Monte Carlo, Pseudorandom Number Generation, Sampling Discrete Random Variables: Inverse Transform Method
- **Week 2:** Discrete: Accept-Reject Algorithm, Composition Method, Sampling Continuous Random Variables: Inverse Transform Method
- **Week 3:** Continuous: Accept-reject Algorithm with examples, Box-Muller method
- **Week 4:** Continuous: Ratio-of-Uniforms method, examples and code, miscellaneous methods in sampling, Sampling from multivariate distributions
- **Week 5:** Simple Importance Sampling: Examples, bias, variance, consistency, Optimal proposals
- **Week 6:** Weighted importance sampling: Examples, Review of likelihood functions, MLE examples
- **Week 7:** Linear regression as MLE, Penalized regression, No-closed form MLEs, Review of Taylor Series Approximations
- **Week 8:** Newton's optimization algorithm: examples and code, Gradient Descent algorithm, applications to logistic regression with code
- **Week 9:** MM algorithm, application to Bridge Regression, EM algorithm, Introduction to Gaussian Mixture Model
- **Week 10:** EM algorithm for GMM, Cross-validation with examples
- **Week 11:** Bootstrapping: examples and code. Application to bridge regression, stochastic gradient descent
- **Week 12:** Applications of SGD with code. Simulated annealing: examples, codes, and challenges

---

## BSCS4021 — Advanced Algorithms

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Neeldhara Misra (Faculty, Dept. of CSE, IIT Gandhinagar)

### Course Description:

To introduce advanced ideas in design of algorithms; To study the performance guarantees of algorithms; To introduce methods for coping with NP-hard problems.

### Syllabus:

- **Week 1:** Greedy Algorithms: Storing Files on Tape; Scheduling Classes; Stable Matchings
- **Week 2:** Matroids: A Generic Optimization Problem, Motivating the Definition, Examples of Matroids, Scheduling with Deadlines
- **Week 3:** Dynamic Programming: Longest Increasing Subsequence, Edit Distance, Subset Sum, Optimal BSTs
- **Week 4:** Maximum Flows: Flows, Cuts, Maxflow-Mincut, Augmenting Paths, Bipartite Matchings, Other Settings
- **Week 5:** Applications of Flows: Exam Scheduling, Baseball Elimination, Project Selection
- **Week 6:** NP-hardness: P, NP, NP-hardness, NP-completeness, Reductions and SAT, 3SAT, Maximum Independent Set, Graph Coloring, Subset Sum

- **Week 7:** Approximation Algorithms: Introduction to Approximation Frameworks, Vertex Cover via Maximal Matchings, Vertex Cover via LP rounding, TSP, Set Cover
- **Week 8:** Randomized Algorithms – Monte Carlo v. Las Vegas, Min-Cut Algorithm, MAX SAT via the Probabilistic Methods, 2SAT via Markov Chains, Primality Testing
- **Week 9:** Exact Algorithms – Branch and Bound, An Inclusion-Exclusion approach to Hamiltonian Path, Dynamic Programming for TSP, Local Search
- **Week 10:** Parameterized Algorithms – Closest String, Iterative Compression for FVS, Randomized Algorithm for k-Path, DP over subsets – Set Cover
- **Week 11:** Kernelization – Vertex Cover, Matrix Rigidity, Feedback Arc Set on Tournaments, Max Sat, Edge Clique Cover
- **Week 12:** Practical Approaches to Coping with Hardness – SAT Solvers, SAT reductions, LP solvers, LP reductions

### BSCS3031 — Computer Systems Design

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Ayon Chakraborty (Dept. of Computer Science and Engineering, IIT Madras)

#### Course Description:

To learn about the internal organization of the computer. To learn about the architecture of a computer's CPU.

#### Syllabus:

- **Week 1:** Course Overview, Building Blocks of Computer Systems
- **Week 2:** Foundation of Logical Circuits, Introduction to Boolean Algebra
- **Week 3:** Canonical Forms (SOP & POS), Universal Gates and Timing Diagrams
- **Week 4:** N-bit Comparator, Multiplexer, Encoder and Decoder
- **Week 5:** Adder and PLA, Boolean Logic Simplification, K-maps
- **Week 6:** Introduction to Circuit verse, 7 Segment LED Display
- **Week 7:** Introduction to Sequential Circuits
- **Week 8:** Registers, Counters
- **Week 9:** Sequential logic design, FSM Design
- **Week 10:** Instruction Set architecture
- **Week 11:** ALU Design
- **Week 12:** CPU Design

### BSCS4022 — Operating Systems

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Chester Rebeiro (Associate Professor, Dept. of Computer Science and Engineering, IIT Madras)

#### Course Description:

Operating systems (OS) provide the crucial interface between a computer's hardware and the applications that run on it. It allows us to write programs without bothering much about the hardware. It also ensures that the computer's resources such as its CPU, hard disk, and memory, are appropriately utilized. In this course, we dwell into how the OS manages to do all this in an efficient manner. This is an introductory course, for students with prior knowledge of computer organization. The course is based on an OS called xv6, which in many ways is similar to the Linux operating systems.

## Syllabus:

- **Week 1:** Introduction and system organization
- **Week 2:** Booting / Memory management
- **Week 3:** Processes and system calls
- **Week 4:** Child process, first process, fork, exit and wait system calls
- **Week 5:** Exec system call, Executables, ELF format
- **Week 6:** Interrupts and interrupts handling
- **Week 7:** Scheduling
- **Week 8:** Synchronization
- **Week 9:** Deadlocks and Threads
- **Week 10:** File system
- **Week 11:** Security: OS security and Side-Channel Attacks
- **Week 12:** Security: TEEs and Buffer Overflow

---

## BSDA5007 — Special Topics in Machine Learning (Reinforcement Learning)

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Balaraman Ravindran (Professor, CSE, IIT Madras)

### Course Description:

To enable the student to understand the reinforcement learning paradigm, to be able to identify when an RL formulation is appropriate, to understand the basic solution approaches in RL, to implement and evaluate various RL algorithms.

### Syllabus:

- **Week 1:** Review of ML fundamentals – Classification, Regression. Review of probability theory and optimization concepts
  - **Week 2:** RL Framework; Supervised learning vs. RL; Explore-Exploit Dilemma; Examples
  - **Week 3:** MAB: Definition, Uses, Algorithms, Contextual Bandits, Transition to full RL, Intro to full RL problem
  - **Week 4:** Intro to MDPs: Definitions, Returns, Value function, Q-function
  - **Week 5:** Bellman Equation, DP, Value Iteration, Policy Iteration, Generalized Policy Iteration
  - **Week 6:** Evaluation and Control: TD learning, SARSA, Q-learning, Monte Carlo, TD Lambda, Eligibility Traces
  - **Week 7:** Maximization-Bias & Representations: Double Q learning, Tabular learning vs. Parameterized, Q-learning with NNs
  - **Week 8:** Function approximation: Semi-gradient methods, SGD, DQNs, Replay Buffer
  - **Week 9:** Policy Gradients: Introduction, Motivation, REINFORCE, PG theorem, Introduction to AC methods
  - **Week 10:** Actor-Critic Methods, Baselines, Advantage AC, A3C; Advanced Value-Based Methods: Double DQN, Prioritized Experience Replay, Dueling Architectures, Expected SARSA
  - **Week 11:** Advanced PG/A-C methods: Deterministic PG and DDPG, Soft Actor-Critic (SAC); HRL: Introduction to hierarchies, types of optimality, SMDPs, Options, HRL algorithms; POMDPs: Intro, Definitions, Belief states, Solution Methods; History-based methods, LSTMs, Q-MDPs, Direct Solutions, PSR
  - **Week 12:** Model-Based RL: Introduction, Motivation, Connections to Planning, Types of MBRL, Benefits, RL with a Learnt Model, Dyna-style models, Latent variable models, Examples, Implicit MBRL. Case study on design of RL solution for real-world problems
-

## BSCS3005 — Programming in C

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Nitin Chandrachoodan (Dept. of Electrical Engineering, IIT Madras)

### Course Description:

This course is intended as a practical introduction to C programming. The focus is on gaining experience with writing and debugging programs. At the end of this course, a student should be able to: write, compile, and run programs in C; use debugging tools to find and correct errors in programs; use various constructs in C and the standard library of C to implement basic data structures and algorithms; understand the need for an OS and how programs interact with the system.

### Syllabus:

- **Week 1:** Write, compile, and run programs in C in a Linux environment; debugging tools
- **Week 2:** Variables, built-in datatypes, operators; Control flow – conditionals, loops
- **Week 3:** Modularity and functions; variable scope
- **Week 4:** Input/Output; Files
- **Week 5:** Pointers, memory, arrays, strings
- **Week 6:** Multi-dimensional arrays, dynamic memory allocation; issues – memory leaks, management
- **Week 7:** Standard library and common extensions (math, time, etc.)
- **Week 8:** Implementation concepts: compilation and execution process; heap/stack; runtime and OS interface

## BSCS4024 — Computer Networks

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Ayon Chakraborty (Assistant Professor, Computer Science and Engineering, IIT Madras)
Pre-requisites	BSCS3005 – Programming in C

### Course Description:

To understand the architecture and functioning of computer networks using a top-down approach. To explore how applications communicate over the Internet and how data is transmitted reliably, efficiently, and securely across layers.

### Syllabus:

- **Week 1:** Introduction to Computer Networking and Internet Architecture
- **Week 2:** Understanding the Layered Protocol Stack and Packet Analysis
- **Week 3:** Fundamentals of Network Performance and Delay Metrics
- **Week 4:** Quality of Experience (QoE) and Web Performance Measurement
- **Week 5:** Foundations of Socket Programming and Client-Server Models
- **Week 6:** Application-Level Protocols and TCP Traffic Analysis
- **Week 7:** Reliable Data Transfer and Transport Layer Mechanisms
- **Week 8:** Flow Control and Sliding Window Protocols in TCP
- **Week 9:** Congestion Control Algorithms and TCP Fairness
- **Week 10:** IP Addressing, Subnetting, and Routing Principles
- **Week 11:** Routing Algorithms and Internetworking
- **Week 12:** Medium Access Control and Wireless Communication Protocols

## BSDA5005 — Introduction to Natural Language Processing (i-NLP)

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Parameswari Krishnamurthy, Rahul Mishra (Assistant Professors, Language Technologies Research Center, IIIT Hyderabad)
Pre-requisites	BSCS3004 – Deep Learning

### Course Description:

Natural language (NL) refers to the language spoken/written by humans. NL is the primary mode of communication for humans. With the growth of the world wide web, data in the form of text has grown exponentially. It calls for the development of algorithms and techniques for processing natural language for the automation and development of intelligent machines: Natural Language Processing (NLP).

### Learning Outcomes:

1. Why is processing language computationally hard and why specialized techniques need to be developed to process texts?
2. Knowledge and in-depth understanding of linguistics techniques and classical (statistical) approaches (pre-deep learning era) to NLP and their limitations
3. Knowledge and in-depth understanding of deep learning approaches (RNN and CNN) to NLP
4. Knowledge and in-depth understanding of Attention Mechanism, Transformers and Large Language Models (LLMs)
5. Ability to read and understand latest NLP-related research papers
6. Ability to identify applicable NLP technique to solve a real-world problem involving text processing
7. Ability to implement NLP models and algorithms for problems related to text processing
8. Ability to develop applications based on textual generative models (LLMs)

### Syllabus:

- Introduction to Natural Language (NL): Why is it hard to process a natural language? Levels of Language Processing, Linguistic Fundamentals for NLP
- Text Processing and Preprocessing: Tokenization, Normalization, Stop word removal, Stemming, lemmatization, Morphological Analysis & Finite State Transducers, Part-of-speech tagging and Named entities
- Classical Sequence Models: HMM and CRF
- Syntax and Parsing: Constituency parsing, Dependency parsing, Parsing algorithms
- Meaning Representation: Distributional Semantics, Logical Semantics, Semantic Role Labelling
- Language Models: n-gram and Word2Vec, GloVe
- Discourse Processing: Anaphora and Coreference Resolution and Discourse Connectives. Machine Translation
- Naïve Bayes Classifier, Expectation Maximization Algorithm, Logistic Regression, Maximum Entropy Models
- Recurrent neural networks, LSTMs/GRUs, Neural Sequence Models, Contextualized Word Embeddings: TagLM, ELMO, ULMFIT, etc., Attention Mechanism (Code demo: LSTM/GRU)
- Transformers, Self-attention Mechanism, Sub-word tokenization, Positional encoding, Pre-trained Language Models (PLMs): BERT, GPT, etc. Fine-tuning and transfer learning (Code demo: Transformers)
- Natural Language Generation, Decoding schemes: greedy, Random sampling, Top-k, Top-p, Speculative sampling, etc., Retrieval Augmented generation (RAG)
- Large Language Models (LLMs): Parameter Efficient Fine Tuning: Prefix-coding, LORA, QLORA, etc. Emergent Behavior: In-context learning, Instruction Fine Tuning, RLHF, DPO (Code demo: LLMs with Hugging face)

- NLP applications: QnA, Summarization, NLI, Fact-checking, etc. Model Explainability: Attention maps, Attention-flow/rollout, Integrated gradients, etc.

## BSDA5006 — Deep Learning for Computer Vision

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Vineeth N B (Dept. of Computer Science and Engineering, IIT Hyderabad)

### Course Description:

To learn about: Knowledge of basics of image processing and computer vision; Knowledge of building blocks of deep learning including feedforward networks, convolutional neural networks, recurrent neural networks and transformers; Knowledge of generative AI models in computer vision; Knowledge of recent trends including explainability/zero-shot learning, few-shot learning, self-supervised learning, etc.; Hands-on experience on implementation of basic image processing tasks; Hands-on experience on implementation of deep learning models for computer vision tasks; Hands-on experience on implementation of advanced computer vision tasks such as explainability, self-supervised learning.

### Syllabus:

- **Week 1:** Introduction and Overview: Course Overview and Motivation; Introduction to Image Formation, Capture and Representation; Linear Filtering, Correlation, Convolution
- **Week 2:** Visual Features and Representations: Edge, Blobs, Corner Detection; Scale Space and Scale Selection; SIFT, SURF; HoG, LBP, etc.
- **Week 3:** Visual Matching: Bag-of-words, VLAD; RANSAC, Hough transform; Pyramid Matching; Optical Flow
- **Week 4:** Deep Learning Review: Review of Deep Learning, Multi-layer Perceptrons, Back propagation
- **Week 5:** Convolutional Neural Networks (CNNs): Introduction to CNNs; Evolution of CNN Architectures: AlexNet, ZFNet, VGG, InceptionNets, ResNets, DenseNets
- **Week 6:** Visualization and Understanding CNNs: Visualization of Kernels; Backprop-to-image/Deconvolution Methods; Deep Dream, Hallucination, Neural Style Transfer; CAM, Grad-CAM, Grad-CAM++; Recent Methods (IG, Segment-IG, SmoothGrad)
- **Week 7:** CNNs for Recognition, Verification, Detection, Segmentation: CNNs for Recognition and Verification (Siamese Networks, Triplet Loss, Contrastive Loss, Ranking Loss); CNNs for Detection: Background of Object Detection, R-CNN, Fast R-CNN, Faster R-CNN, YOLO, SSD, RetinaNet; CNNs for Segmentation: FCN, SegNet, U-Net, Mask-R CNN
- **Week 8:** Recurrent Neural Networks (RNNs): Review of RNNs; CNN + RNN Models for Video Understanding: Spatio-temporal Models, Action/Activity Recognition
- **Week 9:** Attention Models: Introduction to Attention Models in Vision; Vision and Language: Image Captioning, Visual QA, Visual Dialog; Spatial Transformers; Transformer Networks
- **Week 10:** Deep Generative Models: Review of (Popular) Deep Generative Models: GANs, VAEs; Other Generative Models: PixelRNNs, NADE, Normalizing Flows, etc
- **Week 11:** Variants and Applications of Generative Models in Vision: Applications: Image Editing, Inpainting, Superresolution, 3D Object Generation, Security; Variants: CycleGANs, Progressive GANs, StackGANs, Pix2Pix, etc
- **Week 12:** Recent Trends: Zero-shot, One-shot, Few-shot Learning; Self-supervised Learning; Reinforcement Learning in Vision; Other Recent Topics and Applications

## BSDA5004 — Large Language Models

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Prof. Mitesh M. Khapra (Associate Professor, Dept. of Computer Science and Engineering, IIT Kanpur)

### Course Description:

Understanding the Transformer architecture; Understanding the concept of pretraining and fine-tuning language models; Compare and contrast different types of tokenizers like BPE, wordpiece, sentencepiece; Understanding different LLMs architectures: encoder-decoder, encoder-only, decoder-only; Exploring common datasets like C4, mc4, Pile, Stack and so on; Addressing the challenges of applying vanilla attention mechanisms for long range context windows; Apply different types of fine-tuning techniques to fine-tune large language models.

### Syllabus:

- **Week 1:** Transformers: Introduction to transformers – Self-attention – cross-attention – Masked attention – Positional encoding
- **Week 2:** A deep dive into number of parameters, computational complexity and FLOPs – Introduction to language modeling
- **Week 3:** Causal Language Modeling: What is a language model? – Generative Pretrained Transformers (GPT) – Training and inference
- **Week 4:** Masked Language Modeling: Bidirectional Encoder Representations of Transformers (BERT) – Fine-tuning – A deep dive into tokenization: BPE, SentencePiece, wordpiece
- **Week 5:** Bigger Picture: T5, A deep dive into text-to-text (genesis of prompting), taxonomy of models, road ahead
- **Week 6:** Data: Datasets, Pipelines, effectiveness of clean data, Architecture: Types of attention, positional encoding (PE) techniques, scaling techniques
- **Week 7:** Training: Revisiting optimizers, LION vs Adam, Loss functions, Learning schedules, Gradient Clipping, typical failures during training
- **Week 8:** Fine Tuning: Prompt Tuning, Multi-task Fine-tuning, Parametric Efficient Fine-Tuning, Instruction fine-tuning datasets
- **Week 9:** Benchmarks: MMLU, BigBench, HELM, OpenLLM, Evaluation Frameworks
- **Week 10:** Training Large Models: Mixed precision training, Activation checkpointing, 3D parallelism, ZERO, Bloom as a case study
- **Week 11:** Scaling Laws: Chinchilla, Gopher, Palm v2
- **Week 12:** Recent advances

---

## BSMS4023 — Game Theory and Strategy

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Bikramaditya Datta (Assistant Professor, Dept. of Economic Sciences, IIT Kanpur), Vimal Kumar (Professor, Dept. of Economic Science, IIT Kanpur)

### Course Description:

Learn how to think of social and economic aspects of life via mathematical models. Learn how game theory is applied to think about problems in the information economy.

### Syllabus:

- **Week 1:** Introduction and General Principles
- **Week 2:** Games with Simultaneous Moves I

- **Week 3:** Games with Simultaneous Move II
- **Week 4:** Games with Sequential Moves
- **Week 5:** Combining Sequential and Simultaneous Moves and Mixed Strategies
- **Week 6:** Evolutionary game Theory
- **Week 7:** Matching Problem, Gale-Shapley Algorithm
- **Week 8:** Voting, Cascades, and Arrow's Impossibility Theorem
- **Week 9:** Cooperative Games, Shapley Values
- **Week 10:** Fair Division, Bankruptcy Problems
- **Week 11:** Auctions
- **Week 12:** Network Effects

### BSMS3033 — Managerial Economics

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Vimal Kumar (Professor, Dept. of Economic Science, IIT Kanpur)

#### Course Description:

To learn the workings of markets through the prism of Demand and Supply. Understand how a rational consumer makes any choice. Learn the relationships between production, costs, and profits. Learn different forms of markets and their key characteristics. Understand how governments regulate businesses. Learn the business model behind the success of Amazon, UBER, Alibaba, iPhone.

#### Syllabus:

- **Week 1:** Introduction
- **Week 2:** Demand, Supply, and markets
- **Week 3:** Elasticity and Empirical Methods for Demand Analysis
- **Week 4:** Consumer Behavior
- **Week 5:** Technology, Production, and Costs
- **Week 6:** Perfectly Competitive Market
- **Week 7:** Monopoly
- **Week 8:** Monopsony and Monopolistic Competition
- **Week 9:** Decisions under Risk and Uncertainty
- **Week 10:** Asymmetric Information
- **Week 11:** Government and Businesses
- **Week 12:** Platform Businesses

### BSMS3034 — Corporate Finance

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Bikramaditya Datta (Assistant Professor, Dept. of Economic Sciences, IIT Kanpur)

#### Syllabus:

- **Week 1:** Introduction to Finance
- **Week 2:** Present Value, Annuities and Perpetuities
- **Week 3:** Capital Budgeting Rules, Time Varying Rates of Return, Uncertainty, Default and Risk
- **Week 4:** Investment: Risk and Rewards, Choice Under Uncertainty

- **Week 5:** Portfolio Theory
- **Week 6:** Capital Asset Pricing Theory
- **Week 7:** Other Pricing Models: Factor Models, Arbitrage Pricing Theory
- **Week 8:** Market Efficiency: Market Imperfections, and Applications to Capital Budgeting Rules
- **Week 9:** Introduction to Derivative Securities and Basic Options Theory I
- **Week 10:** Introduction to Derivative Securities and Basic Options Theory II
- **Week 11:** Foreign Exchange Markets and International Financial System
- **Week 12:** Financial Crises

### BSDA6003 — Deployability Aspects of AI

Attribute	Details
L-T-P-C	4-0-0-4
Pre-requisite	Introduction to Machine Learning

#### Course Objectives:

The laboratory success of machine learning/deep learning algorithms often do not translate into successful deployment. The aim of this course is to introduce participants to a broad range of issues that crop up when deploying ML algorithms including explainability/fairness/privacy and security related aspects. The focus would be to understand these issues from an algorithmic/statistical viewpoint.

#### Course Contents:

- **Explainability and Interpretability of Black Box ML/DL Models:** Overview of simple interpretable models (linear regression, decision trees); Model agnostic interpretations; example based explanations; Local explanation models; Global Explanation Models; Saliency Maps based Approaches, Parts based Explanations. Theoretical aspects of explainability
- **Causal Machine Learning:** Review of Graphical Models, Structural Causal Models, Pearl's do-calculus, Confounders, Counterfactuals, Causal discovery from data, Intervention based Causal discovery, Causal representation learning
- **Algorithmic Fairness in Machine Learning:** Need for Fair ML, Metrics of fairness – Group fairness: Equality of odds, Demographic parity, Predictive rate parity. Individual fairness, Counterfactual fairness. Impossibility results
- **Private and Federated Machine Learning:** Statistical notions of Private Machine Learning – Differential Privacy. Machine learning in horizontally/vertically partitioned data – statistical/communication complexity/privacy related challenges
- **Transferable Machine Learning:** Transfer Learning – Inductive vs transductive transfer; Active transfer learning; Transferable representation learning. one shot learning. Human in the loop learning
- **Additional Topics:** Meta Learning, Continual Learning, MLOps

#### Text Books:

- Molnar, Christoph. *Interpretable machine learning*. Lulu.com, 2020
- Solon Barocas and Moritz Hardt and Arvind Narayanan, *Fairness and Machine Learning*, Fairmlbook.org, 2019

#### References:

- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017
- Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science* 9.3-4 (2014): 211-407
- Relevant research papers in the topics of interest

## BSDA6001 — Responsible AI

Attribute	Details
L-T-P-C	4-0-0-4
Pre-requisite	Introduction to Machine Learning

### Course Objectives:

AI/ML systems are becoming increasingly pervasive in all facets of our lives. However, deployment of (large scale) AI systems without careful consideration of ethical, legal, privacy, societal concerns might lead to undesirable outcomes. The course is intended to introduce the basic principles of deploying AI in a responsible manner.

### Course Contents:

- **Module 1:** AI Capabilities Improvement in last 5-10 years; Imminent risks from AI Models: Toxicity, bias, goal misspecification, adversarial examples etc.; Long-term risks from AI Models: Misuse, Misgeneralization, Rogue AGI; Principles of RAI – Transparency; Accountability; Safety, Robustness and Reliability; Privacy and Security; Fairness and non-discrimination; Human-Centred Values; Inclusive and Sustainable development, Interpretability; Recap of Deep Learning Techniques, Language/Vision Models; AI Risks for Gen models; Adversarial Attacks – Vision, NLP, Superhuman Go agents
- **Module 2:** ML Poisoning Attacks like Trojans; Implications for current and future AI safety; Explainability; Imminent and Long-term potential for transparency techniques; Mechanistic Interpretability; Representation Engineering, model editing and probing; Critiques of Transparency for AI Safety
- **Module 3:** Privacy & Fairness in AI
- **Module 4:** Metrics and Tools for RAI – measuring bias/fairness, adversarial testing, explanations (Lime/SHAP/GradCam), audit mechanisms; Regulation landscape – DPDP act (India), GDPR (EU), EU AI act, US presidential declaration, Ethical approvals, informed consent, participatory design, future of work, Indian context; What is AGI? When could it be achieved?; Instrumental Convergence: Power Seeking, Deception etc.
- **Module 5:** RAI in Legal, Healthcare, Education domain; Policy issues in RAI

### Text Books:

- Virginia Dignum, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*, Springer Nature, 2019; ISBN-10: 3030303705, ISBN-13: 978-3030303709
- Christoph Molnar. *Interpretable Machine Learning*. Lulu, 1st edition, March 24, 2019; eBook. ISBN-10: 0244768528, ISBN-13: 978-0244768522

---

## BSDA6004 — Sequential Decision Making

Attribute	Details
L-T-P-C	4-0-0-4
Pre-requisite	Introduction to Probability

### Course Objectives:

The course aims to introduce the basic concepts and algorithms related to sequential decision making. Based on the type of feedback received by a system, three different types of sequential learning paradigms will be introduced including online learning (full information), multi-armed bandits (partial information) and reinforcement learning (partial information with state change). Algorithms and applications relevant to each paradigm will be discussed.

### Course Contents:

- **Module 1:** Introduction to Online Learning, Halving algorithm, Online Machine Learning; Perceptron and Winnow, Intro to Regret; Online learning with expert advice – Hedge algorithm; Online linear optimization,

Online convex optimization; Online learning summary

- **Module 2:** Introduction to Multi armed Bandits; Adversarial Bandits – EXP3 algorithm, Contextual MAB – EXP4 algorithm, Stochastic MAB, Epsilon Greedy, Explore then commit, Stochastic MAB, UCB, Thompson Sampling, Stochastic MAB – Linear Bandits LinUCB algorithm; MAB summary
- **Module 3:** Introduction to Reinforcement Learning; Markov Decision Process, Q-learning

**Text Books:**

- Bubeck S. *Introduction to online optimization*. Lecture notes. 2011 Dec 14;2:1-86
- Bubeck, S. and Cesa-Bianchi, N., 2012. *Regret analysis of stochastic and nonstochastic multi-armed bandit problems*. Foundations and Trends® in Machine Learning, 5(1), pp.1-122
- Sutton RS. *Reinforcement learning: an introduction*. A Bradford Book. 2018

**References:**

- Lattimore T, Szepesvári C. *Bandit algorithms*. Cambridge University Press; 2020 Jul 16

**BSDA6005 — Information Theory for Learning**

Attribute	Details
L-T-P-C	4-0-0-4
Pre-requisite	Introduction to Probability

**Course Objectives:**

To introduce information theory as the study of fundamental limits of information processing problems; To study information measures, their properties and applications; To study information-theoretic limits of coding, hypothesis testing and parameter estimation.

**Course Contents:**

- **Information measures:** Entropy, KL divergence, Mutual Information, Total variation, Hellinger distance, f-divergences
- **Coding:** Lossless data compression, Huffman codes, Source coding theorems; Binary Hamming codes, Capacity of binary symmetric channels
- **Hypothesis testing:** Neyman-Pearson lemma, Type I vs Type II error rate regions, Achievability and converse bounds, Asymptotic regimes
- **Parameter estimation:** Statistical decision theory, Minimax and Bayes risk, Upper bounds on minimax risk, Lower bounds on minimax risk – Le Cam, Assouad and Fano methods

**Text Books:**

- *Elements of Information Theory* (second edition) by Thomas M. Cover and Joy A. Thomas
- *Information Theory: From coding to learning* by Yury Polyanskiy and Yihong Wu

**BSDA6002 — Statistical Learning Theory**

Attribute	Details
L-T-P-C	4-0-0-4
Pre-requisites	Introduction to Probability, Introduction to Machine Learning

**Course Objectives:**

The goal of the course is to introduce statistical notions of learning from data and various models including the PAC model for learning. The course will cover various classical notions of learning including uniform convergence based results of Vapnik to more recent ideas of algorithmic stability. The course will also give a

broad introduction to the theory of unsupervised learning and a high level overview of theory of modern deep learning.

**Course Contents:**

- **Module 1:** Preliminaries, Probability review, Concentration inequalities. Machine Learning review, PAC Learning, Generalization bounds: Vapnik-Chervonenskis dimension
- **Module 2:** Data Dependent Generalization bounds – Rademacher Complexity, Algorithmic Stability, Statistical Consistency Analysis, Structural Risk Minimization
- **Module 3:** Theory of Unsupervised Learning
- **Module 4:** Theory of Modern Deep Learning: Infinite width neural networks, Neural tangent kernels, inductive bias due to algorithmic regularization

**Text Books:**

- L. Devroye, L. Györfi, G. Lugosi. *A probabilistic theory of Pattern Recognition*. Springer, 1996
- Vapnik V. *The nature of statistical learning theory*. Springer science & business media; 2013 Jun 29

**References:**

- Cesa-Bianchi, and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006
- Roberts DA, Yaida S, Hanin B. *The principles of deep learning theory*. Cambridge, MA, USA: Cambridge University Press; 2022 May 5
- Ghahramani (2004) Unsupervised Learning. In Bousquet, O., Raetsch, G. and von Luxburg, U. (eds) *Advanced Lectures on Machine Learning* LNAI 3176. Springer-Verlag

**BSDA5014 — ML Ops**

Attribute	Details
L-T-P-C	4-0-0-4
Instructor(s)	Dr. Rangarajan Vasudevan (Co-Founder & Chief Data Officer, Lentre.ai)
Pre-requisites	BSCS1002 – Programming in Python, BSCS2008 – Machine Learning Practice

**Course Contents:**

- **Week 1:** Introduction to MLOps: Overview of MLOps and its significance; Key challenges in deploying and managing ML models in production, Comparison of traditional software development, DevOps and MLOps, Key components of MLOps, MLOps workflow, Landscape of MLOps tools and technologies
- **Week 2:** ML Pipelines & Data Management: Overview of data engineering tools and practices, Data management for ML models, ML pipeline automation. DVC overview
- **Week 3:** Data Management – Part 2: Feature Stores. Motivation, role in ML and Generative AI applications, benefits for MLOps. Feast overview
- **Week 4:** CI/CD for ML Models: Use of version control systems like Git for model development, automated testing & validation, model delivery strategies
- **Week 5:** Machine Learning Model Development: Comparison of development of small models vs large models including LLMs, tracking model training & experimentation using MLFlow, hyperparameter tuning and optimization
- **Week 6:** Model Deployment and Serving: Overview of containerization and orchestration technologies and various other cloud-based deployment options, Deployment strategies for different environments (cloud, edge, on-premises), Model serving options
- **Week 7:** Monitoring and Performance Optimization: Techniques for monitoring model performance in production, Logging and error tracking for ML systems, Performance optimization and scaling strategies
- **Week 8:** ML Security: Overview of Security considerations for ML, field of MLSecOps, tooling options

- **Week 9:** ML Governance: Overview of model explainability and ethical considerations in ML deployments, tracking bias
- **Week 10:** MLOps for LLMs – Part 1: Model Versioning for base models and fine-tuned variants, CI/CD specifics
- **Week 11:** MLOps for LLMs – Part 2: Accuracy, Performance vs Cost tradeoffs, Observability, Security & Governance (Bias, Toxicity, Explainability)
- **Week 12:** Closing topics: Advanced topics of Federated learning, edge inferencing; Putting it all together

**Prescribed Books:**

- *Building Machine Learning Powered Applications: Going from Idea to Product* – Emmanuel Ameisen – O'Reilly publication
- *Reliable Machine Learning: Applying SRE Principles to ML in Production* – Chen, Murphy, Parisa – O'Reilly publication
- *Machine Learning Engineering in Action* – Ben Wilson – O'Reilly publication
- *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems* – Martin Kleppmann – O'Reilly publication

**BSDA5002 — Mathematical Foundations of Generative AI**

Attribute	Details
L-T-P-C	4-0-0-4
Pre-requisite	Introduction to Machine Learning

**Course Objectives:**

The course will cover an introduction to generative AI focussing on models used in computer vision and large language models along with selected applications.

**Course Contents:**

- **Module 1: Introduction (2 weeks):** Machine learning recap, generative vs discriminative modeling, introduction to classical generative models – Naive Bayes; Structured vs unstructured data, artificial neural network, introduction to deep learning
- **Module 2: Models and applications (6-8 weeks):** Representation learning – PCA recap, Auto encoders, variational autoencoders, Generative Adversarial networks and variants, Auto regressive models – LSTM and variants, Flow based Models – Normalized flow models, Energy based Models, Diffusion Models, Transformer based Models. Applications: Conversational AI – general and domain specific (biology, education, etc), learning to generate images, learning to generate music, learning to write computer programs. Multimodal generative AI, towards AGI
- **Module 3: Reinforcement Learning for Generative AI (2 weeks):** RLHF – RL from human feedback
- **Module 4: Pitfalls (1 week):** Data bias, hallucination, privacy violations

**Text Book:**

- Foster D. *Generative deep learning*. O'Reilly Media, Inc.; 2023

**References:**

- Recent papers/surveys that are relevant to the course

## BSDA5003 — Algorithms for Data Science

Attribute	Details
L-T-P-C	4-0-0-4
Pre-requisites	Introduction to Probability, Introduction to Linear Algebra

### Course Objectives:

The aim of this second level, one semester graduate course is to provide a broad overview and develop the tools and methods necessary for the large-scale problems that naturally arise in many application areas.

### Course Contents:

- **Introduction and Overview:** Evolution of data science over centuries. The three pillars in the development of predictive science – Building models, estimation of parameters of the model and prediction. A classification of problems – direct and inverse problems. Uni vs. multivariate data analysis. Sources of big data and a classification of data sets. Relation between spatial data analysis and big data Analytics. Role of high dimensional Statistics and Probability theory
- **Model building:** Multivariate regression analysis and principles of least squares estimation and prediction. Best linear unbiased estimators (BLUE). Gauss-Markov theorem. Sensitivity Analysis an outlier deduction, Subset selection to improve the quality of forecast. Ridge regression, Sparse modeling – LASSO and its variants
- **Dimension Reduction:** Theory of principal component analysis, its inherent optimality, The notion of empirical orthogonal functions and their applications to analysis of geophysical data
- **Curse of dimensionality and its manifestations:** Volume of spheres in high dimension and creation of empty space, relation between volumes of cubes and spheres in high dimension, Concentration of distances, Concentration of probability measures and concentration of functions in high dimensional spaces and their impact
- **Randomized Algorithms in Linear Algebra – I:** Sampling based approach. Two types of sampling – uniform and importance sampling. Concept of the sketch of a matrix. Approximating a large matrix by a low rank counterpart, Approximation of product of two large matrices – optimal sampling probabilities. Quantifying the quality of approximation using error bounds using Frobenius and spectral norms. Uncertainty principle of Fourier transform. Reduction of time complexity using randomized Hadamard transform and uniform sampling. Application to solving large scale least squares problems
- **Randomized Algorithms in Linear Algebra – II:** A discussion of the need for and classification of methods for dimension reduction. Isometry and near isometry. Orthogonal projections, Concentration Inequalities, Dimension reduction using random projections, Johnson-Lindenstrauss Lemma, and various strategies for implementation. Applications to solving large scale least squares problems, compressed sensing, Image analysis
- **Hashing:** Locality Sensitive Hashing – introduction and applications, bloom filters, similarity search

### Text Books:

- A. Blum, J. Hopcroft, and R. Kannan (2020) *Foundations of Data Sciences*, Cambridge University Press
- M. W. Mahoney (2010) *Randomized Algorithms for Matrix and Data*, Foundations and Trends in Machine Learning, pages 123-224

### References:

- J. M. Lewis, S. Lakshminarayanan and S. K. Dhall (2006) *Dynamic Data Assimilation: a least squares approach*, Cambridge University Press
- V. Shikhman and D. Muller (2021) *Mathematical Foundations of Big Data Analytics*, Springer
- I. Rish and G. Ya. Grabarnik (2015) *Sparse Modeling; Theory, Algorithms and Applications*, CRC Press
- M Wainwright (2019) *High-Dimensional Statistics: A non-asymptotic viewpoint*, Cambridge University Press
- J. Wang (2012) *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*, Springer

- N. Halko, P.G. Martinsson, and J. A. Tropp (2011) *Finding Structures with Randomness: Probabilistic Algorithms for constructing Approximate Matrix decompositions*, SIAM Review, Vol 53, 217-288
- D. P. Woodruff (2015) *Sketching as tool for linear algebra*, David P. Woodruff, Sketching as a Tool for Numerical Linear Algebra, now, 2014, doi: 10.1561/04000000060

## BSDA5013 — Deep Learning Practice

Attribute	Details
<b>L-T-P-C</b>	4-0-0-4
<b>Instructor(s)</b>	Prof. Mitesh Khapra (Dept. of Computer Science and Engineering, IIT Madras), Prof. Umesh (Dept. of Electrical Engineering, IIT Madras), Dr. Kaushik Mitra (Assistant Professor, Dept. of Electrical Engineering, IIT Madras)

### Syllabus:

- **Week 1: Introduction to Modern NLP and HF**

Introduction to Modern NLP with transformers. Introduction to Hugging Face Ecosystem: Datasets, Tokenizer, Transformers, Trainers, Accelerate (slides). Dive into the datasets module (code). Loading datasets from the hub, local files, understanding the base class. Some frequently used methods: map, filter, select, flatten, concatenate, interleave. Load large datasets by streaming. (optional) Building datasets by scraping web (slides/code)

- **Week 2: Tokenization**

Understand the working principle of different tokenization algorithms (slides). Tokenizer pipeline: Normalization, Pre-tokenization, Tokenization algorithm, Post-processing. Training a tokenizer to learn the vocabulary (code). Encoding and decoding of samples (code)

- **Week 3: Fine-tuning models for downstream tasks**

Use pre-trained models for various NLP tasks. Fine-tune models on the task-specific datasets. Freeze the pre-trained model parameters and only fine-tune the task-specific head. Full parameter fine-tuning. Performance evaluation using appropriate metrics

- **Week 4: Continual Pre-training and Instruction Tuning**

Train a model from scratch using CLM and MLM objectives. Memory efficient training: Optimizers: AdaFactor, 8-bit Adam, (Optional) GaLore; Gradient Accumulation; Activation Checkpointing; Mixed Precision Training. Continual pre-training using PEFT (LORA, QLORA). (optional) Instruction tuning

- **Week 5: Spoken Language Identification**

- **Week 6: Speaker Diarisation (identifying speakers and when they are spoken in a conversation)**

- **Week 7: Speech to Text and Text to Speech synthesis**

- **Week 8: Wake word detection like "Hey Google" or "Alexa" with personalization**

- **Week 9: Image Classification**

Models: Alexnet, VggNet, ResNet. Explore: densenet, ViT. Dataset: 10 class animal dataset. Challenges: large within variation, so difficult to classify. Can introduce class imbalance

- **Week 10: Object detection**

Models: YOLO, RCNN. Dataset: Mosquito detection and recognition. Challenge: No pretrained model, class imbalance

- **Week 11: Image based depth estimation**

Dataset: Challenges: low-light, so directly cannot use pre-trained models. Models: UNet, UNet++, Pix2Pix

- **Week 12: Image super-resolution**

Dataset: Medical image dataset. Challenges: no pretrained model. Models: SRGAN

## BSCS3021 — Theory of Computation

Attribute	Details
Credits	4
Course Type	Elective
Prerequisite	None
L-T-P-C	4-0-0-4

### Course Description:

Theory of computation deals with the encapsulation and abstraction of diverse computational processes, whether hardware or software, which enables us to compare them and understand their basic capabilities and limitations. It is intrinsically related to algorithmic models, where we seek to determine what can and cannot be computed, how quickly, with how much memory, without being bogged down by low-level implementation details. The course will introduce the tools for using and arguing about various fundamental computational questions in a rigorous framework. It will elucidate how formal languages help understanding of core techniques like simulation, reductions, and nondeterminism. Together, it will offer deep understanding of what is hard to compute efficiently and trade-offs between resources like time and space even for the fastest computing machines.

### Learning Outcomes:

In this course a student will:

- Understand formal models of computation including finite-state machine and pushdown automata and associated languages
- Understand the concept of non-determinism
- Design and simulate Turing Machines to explore the boundaries of computability
- Analyze computational complexity
- Use reductions to relate problems to argue about computability, NP-Completeness and the limits of efficient computation
- Appreciate the landscape of complexity theory
- Develop formal reasoning and proof skills

### Week-wise Syllabus:

- **Week 1:** Introduction to Theory of Computation with Finite Automata
  - **Week 2:** Regular Languages and Regular Expressions
  - **Week 3:** Regular Languages, DFA Minimization, and Pumping Lemma
  - **Week 4:** Context-Free Languages (CFLs) and Parse Trees
  - **Week 5:** Pushdown Automata and Recognizers of CFLs
  - **Week 6:** Non-CFLs and Introduction to Turing Machines
  - **Week 7:** Variants of Turing Machines and the Church–Turing Thesis
  - **Week 8:** Decidability and Undecidable Problems
  - **Week 9:** Reductions and Proving Undecidability
  - **Week 10:** Time and Space Complexity
  - **Week 11:** Polynomial-Time Reductions and Hard Problems
  - **Week 12:** NP-Completeness and Course Summary
-

## BSDA4001 — Data Science and AI Lab

Attribute	Details
Credits	4
Course Type	Elective
Prerequisite	Deep Learning
L-T-P-C	4-0-0-4

### Course Description:

This is a hands-on course designed to bridge the gap between academic knowledge and industry needs in Data Science and AI. The course revisits essential data science concepts and includes contemporary AI technologies. Each week introduces a new concept and accompanies an end-to-end assignment, simulating real-world workflows.

### Syllabus:

- **Week 1: Data Science & Python Stack Refresher**  
Key tools: NumPy, Pandas, Matplotlib, Seaborn, Jupyter, Git
  - **Week 2: Machine Learning with Scikit-learn**  
Core models: regression, classification, clustering. Pipelines, hyperparameter tuning, evaluation metrics
  - **Week 3: Deep Learning with PyTorch and TensorFlow**  
Building and training deep neural networks. Model saving/loading, GPU training
  - **Week 4: Computer Vision and Image Processing**  
Using OpenCV, PIL, PyTorch/TensorFlow for image loading, preprocessing. Transfer learning with CNNs (ResNet, MobileNet)
  - **Week 5: NLP and LLMs using HuggingFace**  
Using BERT, RoBERTa for classification/generation. Tokenization, fine-tuning, and inference
  - **Week 6: Prompt Engineering, LangChain & LLM APIs**  
Working with LangChain, templates, and memory modules. Prompt tuning best practices. Using OpenAI, Gemini, and Ollama APIs for inference
  - **Week 7: Retrieval-Augmented Generation (RAG) & Vector DBs**  
Concepts of semantic search & embedding spaces. Use of FAISS, Pinecone, ChromaDB
  - **Week 8: Agents and Autonomous AI (Langchain, LangGraph)**  
Concepts of tool use, planning, multi-step task agents. Implementing basic agents using LangChain
  - **Week 9: Data Visualization & Dashboards**  
Tools: PowerBI, Tableau, Streamlit, Plotly. Dashboard design, storytelling with data
  - **Week 10: Time Series Forecasting**  
Exploratory Analysis of Time Series. Forecasting using Facebook Prophet and DeepAR. Trend, seasonality, holidays, anomaly detection
  - **Week 11: Big Data and Cloud Platforms – Google Cloud Platform**  
Build and run an end-to-end pipeline in GCP. Working with GCP, BigQuery, Vertex AI. Concepts of Spark, Hadoop, Airflow, and model deployment
  - **Week 12: Model Explainability, Ethics & Mini-project**  
Tools: SHAP, LIME for understanding predictions. Concepts: data privacy, fairness, AI ethics. Mini-project
-

Attribute	Details
Course Credits	2
Course Type	Programming – Degree Level Elective
L-T-P-C	4-0-0-4

**Course Content:**

This course is designed to equip students with the essential skills required for web application development roles. By the end of the course, students will be proficient in both frontend and backend technologies, understand version control best practices, and be capable of building real-time web applications. The curriculum focuses on hands-on learning, industry-relevant tools, and best coding practices to make students job-ready.

**Syllabus:**

- **Week 1: Course Overview & Practice DSA**

Overview of Course; Setting Expectations & Goals; Efficiency & Complexity Analysis; Data Structures: Arrays & Strings, Hash Tables & Sets, Linked Lists, Stacks & Queues, Heap, Trees & Graphs, Union-Find data structures; Algorithm Design Techniques: Recursion & Backtracking, Divide and Conquer, Greedy Algorithms, Dynamic Programming; Problem-Solving

- **Week 2: JavaScript Essentials for Web Development**

The role of JavaScript in modern web apps; ES6+ Features: let/const, arrow functions, template literals, destructuring; Functions, Arrays, Objects; Asynchronous JS: Callbacks, Promises, async/await; Hands-on: Interactive JS challenges and code snippets

- **Week 3: Developer Tools & Workflows**

How the web works: Client-Server, HTTP, APIs; Introduction to Git and GitHub/GitLab; Version Control: init, add, commit, push, pull, branching; npm, package.json; Hands-on: GitHub setup and npm usage in a mini project

- **Week 4: Introduction to Backend Development with Node.js & TypeScript (Part 1)**

What is Node.js? Event-driven architecture & single-threaded nature; Node.js in the web development ecosystem; Setting Up a TypeScript Project with Node.js; tsconfig.json basics; Compiling TypeScript and running with ts-node; Modules, imports/exports (ES6-style); Writing a basic HTTP server in TypeScript; Hands-on: simple route handling

- **Week 5: Building REST APIs with Express.js & TypeScript (Part 2)**

Introduction to Express.js with TypeScript; Defining routes, handling requests/responses; Middleware concepts and implementation; Structuring an Express app in TypeScript; Error handling & basic input validation; Hands-on: Building a small CRUD API (e.g., Task Manager or Notes API)

- **Week 6: SQL & PostgreSQL**

What is SQL? Relational DB concepts: Tables, Keys; Learning to write basic queries in SQL; Joins in SQL; Data Filtering using SQL; Pagination in SQL

- **Week 7: Integrating Express.js with PostgreSQL**

Introduction to ORM (TypeORM or Prisma); Entity design & schema modeling; Connecting Express to PostgreSQL; Hands-on: Full CRUD API with DB persistence

- **Week 8: Authentication & Application Security**

Introduction to Authentication; Hashing passwords with bcrypt; JWT-based authentication; Route protection and user sessions; Hands-on: Secure login & register flow

- **Week 9: React.js Fundamentals – Components & JSX**

What is React? Understanding Virtual DOM; Setting up a React Project with Vite; JSX & Component-Based Architecture; Functional Components & Props; Hands-on: Building Simple UI Components

- **Week 10: React.js – State Management & Event Handling**

Introduction to State (useState Hook); Handling User Input & Events; Conditional Rendering; Lists & Keys in React; useEffect Hook; Hands-on: Building Interactive Components

- **Week 11: React.js – Routing & API Integration**

Introduction to React Router; Implementing Navigation & Dynamic Routing; Making API Calls using Fetch / Axios; Handling API Responses & Errors (+ CORS); Hands-on: Fetching & Displaying Data in a React App

- **Week 12: Full-Stack Project – Planning & Development**

Structuring a Full-Stack Application; Combining React.js, Express.js, and Databases; Implementing CRUD Operations End-to-End; Hands-on: Developing a Full-Stack Feature; Deployment – which platform?

---

## BSDA6901 — M.Tech Project

Attribute	Details
No. of Credits	20

### Project:

The project will be in the broad area of machine learning and AI. It could be done in a company or a research lab. It will be executed and evaluated similarly to web MTech programs. Students can start the project after completing the core courses of the PG Diploma.

---

## Credit Transfer & Additional Information

### Credit Transfer

- Apprenticeship courses available:
  - BSCS4005 – Apprenticeship in Data Science 1
  - BSCS4006 – Apprenticeship in Data Science 2
  - BSCS4007 – Apprenticeship in Programming 1
  - BSCS4008 – Apprenticeship in Programming 2

### Comprehensive Exam Courses

Two new 2-credit courses (Level 4) introduced as Electives in the degree level:

Course Code	Course Title	L-T-P-C
<b>BSDA4002</b>	Comprehensive Exam – Data Science & Artificial Intelligence (BD)	4-0-0-4
<b>BSCS4009</b>	Comprehensive Exam – Computer Science & Information Technology (BP)	4-0-0-4

**Note:** Those who write GATE and come in Top 1 percent of all students who appeared in GATE for the respective paper can get 2 credits transferred in the Degree Level. GATE Score will be valid for 2 years and these courses can be credited both in BSc and BS level also.

### NPTEL Credit Transfer

#### SWAYAM NPTEL Approved Dep/Free Elective course list:

[View List](#)

#### SWAYAM NPTEL Approved HS/MG course list:

[View List](#)

**Note:** The list of courses is subject to periodic revision. From the NPTEL course URL, please select the **SWAYAM Certification courses**. Only these are eligible for Credit Transfer.

---

Document formatted in Markdown for clarity and easy navigation. All content sourced from the official IIT Madras BS Degree in Data Science and Applications syllabus.