COEP Technological University Pune
(A Unitary Public University of Govt. of Maharashtra)

NEP 2020 Compliant

Proposed Curriculum Structure

M. Tech. Data Science

(Effective from: A.Y. 2024-25)

List of Abbreviations

Abbreviation	Title	No of courses	Credits	% of Credits
PSMC	Program Specific Mathematics Course	1	4	5.88%
PSBC	Program Specific Bridge Course	1	3	4.41%
PCC	Program Core Course	6	18	26.47%
PEC	Program Specific Elective Course	3	9	13.24%
LC	Laboratory Course	5	5	7.36%
VSEC	Vocational and Skill Enhancement Course	2	18	26.47%
OE	Open Elective	1	3	4.41%
SLC	Self Learning Course	2	6	8.82%
AEC	Ability Enhancement Course	1	1	1.47%
MLC	Mandatory Learning Course	2	--	--
CCA	Co-curricular and Extracurricular Activities	1	1	1.14%
	Total	**25**	**68**	**100%**

Semester I

Sr. No.	Course Type	Course Code	Course Name	Teaching Scheme				Credits
				L	T	P	S	
1	PSMC		Probability and Statistics foundation	3	1	0	1	4
2	PSBC		Algorithms and Complexity Theory	3	0	0	1	3
3	PCC & LC		SQL and Python Programming	3	0	2	1	4
4	PCC & LC		Machine Learning	3	0	2	1	4
5	PCC		Data Engineering	3	0	0	1	3
6	AEC		Mini Project/Seminar	0	0	2	1	1
7	PEC		Program Specific Elective Course-I 1. Foundations of Statistics and Linear Algebra 2. Data Visualization with Tableau 3. Artificial Intelligence 4. Industry offered Electives	3	0	0	1	3
8	MLC		Research Methodology and Intellectual Property Rights	0	0	0	2	-
9	MLC		Effective Technical Communication Skills	0	0	0	1	-
				Total Credits				**22**

Semester II

Sr. No.	Course Type	Course Code	Course Name	Teaching Scheme				Credits
				L	T	P	S	
1	OE		Open Elective: Data Structures	3	0	0	1	3
2	PCC & LC		Big Data Analytics with Apache Spark	3	0	2	1	4
3	PCC & LC		Advanced Machine Learning and Deep Learning	3	0	2	1	4
4	PCC &LC		ML Ops and Systems	3	0	2	1	4
5	PEC		Program Specific Elective –II 1. Time Series Data Analysis 2. Computer Vision 3. R Programming	3	0	0	1	3
6	PEC		Program Specific Elective –III 1. Generative Adversarial Networks 2. Reinforcement Learning 3. Natural Language Processing 4. Industry offered Electives	3	0	0	1	3
7	CCA		Liberal Learning Course	0	0	2	2	1
			Total Credits					**22**

1. **Exit option to qualify for PG Diploma in Data Science :**
   - Eight weeks domain specific industrial internship in the month of June-July after successfully completing first year of the program

Semester III

Sr. No.	Course Type	Course Code	Course Name	Teaching Scheme				Credits
				L	T	P	S	
1	SLC		Massive Open Online Course –I	3	--	--	-	3
2	SLC		Massive Open Online Course –II	3	--	--	-	3
3	VSEC		Dissertation Phase – I	--	--	12	18	6
				Total Credits				**12**

Semester IV

Sr. No.	Course Type	Course Code	Course Name	Teaching Scheme				Credits
				L	T	P	S	
1	VSEC		Dissertation Phase – II	--	--	24	12	12
				Total Credits				**12**

Probability and Statistics foundation ()

Teaching Scheme
Lectures: 3 hrs/week
Tutorial: 1hr/week

Examination Scheme
T1, T2 – 20 marks each
End-Sem Exam – 60

Course Outcomes:
Students will be able to:
1. Solve various problems on probability and statistics.
2. Analyze the given probabilistic model of the problem.
3. Use the techniques studied in probability and statistics to solve problems in domains such as data mining, machine learning, network analysis.

Unit 1: Basic Probability Theory   **[2 Hrs]**
Probability axioms, conditional probability, independence of events, Bayes' rule, Bernoulli trials

Unit 2: Random Variables and Expectation   **[10 Hrs]**
Discrete random variables: Random variables and their event spaces, Probability Mass Function, Discrete Distributions such as Binomial, Poisson, Geometric etc., Indicator random variables, Continuous random variables: Distributions such as Exponential, Erlang, Gamma, Normal etc., Functions of a random variable, Expectation: Moments, Expectation based on multiple random variables, transform methods, Moments and Transforms of some distributions such as Binomial, Geometric, Poisson, Gamma, Normal

Unit 3: Stochastic Processes   **[6 Hrs]**
Introduction and classification of stochastic processes, Bernoulli process, Poisson process, Renewal processes

Unit 4: Markov chains   **[8 Hrs]**
Discrete-Time Markov chains: computation of n-step transition probabilities, state classification and limiting probabilities, distribution of time between time changes, M/G/1 queuing system. Continuous-Time Markov chains: Birth-Death process (M/M/1 and M/M/m queues), non-birth-death processes, Petri nets

Unit 5: Statistical Inference   **[8 Hrs]**
Parameter Estimation – sampling from normal distribution, exponential distribution, estimation related to Markov chains, Hypothesis testing.

Unit 6: Regression and Analysis of Variance   **[6 Hrs]**
Least square curve fitting, Linear and non-linear regression, Analysis of variance.

Text Books:
1. Ronald Walpole, Probability and Statistics for Engineers and Scientists, Pearson, ISBN-13: 978-0321629111

References:
1. Kishor Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, John Wiley and Sons, New York, 2001, ISBN number 0-471-33341-7

Algorithms and Complexity Theory ()

Teaching Scheme: **Examination Scheme:**
Lectures: 3 hrs/week   T1, T2 – 20 marks each
                                                                     End-Sem Exam – 60

Course Outcomes:
Students will be able to:
1. Analyze the asymptotic performance of algorithms
2. Apply important algorithmic design paradigms for solving problems
3. Understand the complexity theory

Unit-I: Mathematical Foundation   **[6 Hrs]**
Growth of functions – Asymptotic notation, Standard notation and common functions,
Summations, solving recurrences.

Unit-II: Analysis of Algorithms   **[8 Hrs]**
Necessity of time and space analysis of algorithms, Worst case analysis of common algorithms
and operations on elementary data structures (e.g. Heapsort), Average case analysis of Quicksort,
Amortized analysis.

Unit-III: Standard Design Techniques-I   **[6 Hrs]**
Divide and Conquer, Greedy method.

Unit-IV: Standard Design Techniques-II   **[8 Hrs]**
Dynamic programming, Graphs and Traversals.

Unit-V: Standard Design Techniques-III   **[6 Hrs]**
Backtracking, Branch-and-bound.

Unit VI: Complexity Theory   **[6 Hrs]**
Lower-bound arguments, Introduction to NP-Completeness, Reducibility (SAT, Independent Set,
3VC, Subset Su, Hamiltonian Circuit etc), Introduction to approximation algorithms

Text Books:
1. Thomas Cormen, Charles Leiserson, Ronald Rivest and Cliford Stein, "Introduction to Algorithms" , PHI

Reference Books:
1. Horowitz and S. Sahni. "Fundamentals of Computer Algorithms", Galgotia, 1991

SQL and Python Programming

Teaching Scheme

Lectures: 3 hrs/week

Lab : 2 hrs/week

Examination Scheme

T1, T2 – 20 marks each

End-Sem Exam – 60 marks

Course Outcomes:

Students will be able to:

1. Design databases and Implement DDL, DML for relational databases.
2. Use aggregate functions to fire queries on the databases.
3. Implement programs using Python.
4. Visualize data and perform Exploratory Data Analysis using Python

SQL

Database Design: Database System Applications, Purpose of Database, Database Architectures, Database Properties, Database Languages, Views of Data, Instances and Schema.

Data Modelling: Relational Model, Data Modelling: ER Model, Attributes and their types, Relationships, Cardinalities, Extended ER Diagram, Specialization and Generalization, Aggregation and Attribute Inheritance.

Relational Databases: Relational Model, Keys concepts, Integrity Constraints, Introduction to SQL: Data Types and Literals, DDL, DML statements, Views: Creating, Dropping, Updating using Views, Indexes, Handling Nulls. Basic Filtering and Advance Filtering with SQL, Wildcards in SQL, Sorting and Math Operations, Built-in Database

Functions: Numeric, Set Operations.

Advanced SQL: Aggregate Functions, Grouping Data with SQL, Sorting and Ordering Data with SQL, SQL Join Operations, Sub Queries, SQL Join Operations, Sub Queries, working with text and Strings, working with Date and Time Strings, Windows functions, Accessing database with Python, Analysing Data with Python, Working with Real world Dataset.

Python

Basic of Python Programming : Introduction to Python and Jupyter Notebooks, Data Types and Typecasting, ipython magic commands and shell commands Control Structures, User Input, Exception Handling, Operators in python.

Data Structures and Functions: Strings, Indexing and Slicing, String Operators, Useful methods for string manipulation and processing, Escape Sequences, String Formatting. List, inbuilt functions with list, in and not in operators, Tuples, difference between tuples and list. Dictionary, dictionary methods, creating dictionary from text, sorting dictionary, inverting dictionary, writing functions in Python, function design recipe, calling functions within another functions.

List and Dictionary Comprehensions: List comprehension, List comprehension using for, nested for loops, if, if-else. Dictionary comprehension, Useful Functions: Lambda operator for defining anonymous functions, Map, filter and reduce functions. Enumerate and zip functions.

Numpy: Creating ndarray, Array indexing and slicing, Integer Indexing, Boolean Indexing, Broadcasting, Arithmetic and Statistical operations, Any and all conditionals, Merging and splitting, Important built-in methods (np.zero, np.ones, np.where, np.unique etc …..).

Data Manipulation and Analysis using Panda: Creating series, Data manipulation with series, Creating Dataframes, Data manipulation with data frames, Data Cleaning, Data Analyzing, Data Visualization, Exploratory Data Analysis using pandas, importing data in python.

Suggested List of Assignments for SQL:

1. Create a schema using SQL DDL Commands and perform basic SQL DML commands on that schema .
2. Write SQL queries using clauses such as Group by, Order by, Having etc. for statements given.
3. Write SQL queries using aggregate functions for given statements on a given schema.
4. Write Nested Subqueries inSQL using Set membership and Set Comparison Commands such for a given statement on a given set of relations.
5. Write SQL queries using join (left, right and full outer join) operations for the statements given.
6. Write SQL queries using Date Functions, String Functions and Math Functions for statements given.
7. Write SQL queries to create a view, drop a view and select from a view for a given statement on a given schema.
8. Create the schema and specify constraints on the given relations using statements given.
9. Demonstrate database connectivity in Python.
10. Draw an E-R diagram and convert entities and relationships to a relation table for a given scenario. Normalize the database up to appropriate normal form.

Suggested List of Assignments for Python :
1. Getting started with Jupyter Notebook, Implementing basic programming concepts.
2. Implementing Data structures and their functionalities.
3. Applying list comprehension and use lambda operators with map, filter and reduce functions to solve a scenario.
4. Analyzing Image Data using Numpy.
5. Perform Data Manipulation on a dataset using Panda.
6. Perform EDA (Exploratory Data Analysis) on a dataset using Panda.
7. Perform Data visualization using static and interactive plots on a given dataset.
8. Connect to a RDBMS and execute SQL queries on a dataset.

Useful Links:
1. https://www.w3schools.in/dbms
2. https://www.w3schools.com/sql

Books:
1. Abraham Silberschatz, Henry F. Korth, S. Sudarshan, "Database system concepts", 5th Edition, McGraw Hill InternationalEdition.
2. Raghu Ramkrishnan, Johannes Gehrke, "Database Management Systems", Second Edition, McGraw Hill InternationalEditions
3. RamezElmasri and Shamkant B. Navathe, "Fundamental Database Systems", 3rdEdition, Pearson Education,2003
4. Rob Coronel, "Database systems : Design implementation and management", 4thEdition, Thomson Learning Press.
5. Richard L. Halterman, Learning to Program with Python.
6. Miller Curtis, Hands-On Data Analysis with NumPy and pandas, Packt Publishing Limited, ISBN: 9781789530797.
7. Jake Vanderplas, Python Data Science Handbook, O'Reilly Media, ISBN: 978-1-491-91205-8

<center>**Machine Learning**</center>

Teaching Scheme   **Examination Scheme**
Lectures: 3 hrs/week   T1, T2 – 20 marks each
Lab: 2 hrs/week   End-Sem Exam – 60 marks

Course Outcomes:
Students will be able to:
1. Understand ML types, challenges, and learning systems.
2. Apply regression techniques and SVM. Naive Bayes, KNN algorithms for Classification.
3. Develop and optimize decision tree models for regression/classification.
4. Utilize Advanced Techniques in Data Analysis for feature reduction.

Contents:
Unit 1: Introduction to Machine Learning: **[6 Hrs]**

Machine Learning applications, Types of ML supervised ML, unsupervised ML, semi-supervised ML, reinforcement ML, batch and online ML systems, instance based and model-based ML systems, Challenges in ML, Overfitting and underfitting.

Unit 2: Supervised Learning: **[6 Hrs]**

Linear Regression, Best fit regression line using least square and gradient descent method, performance metrics for regression, Logistic Regression.

Unit 3: Classification Algorithms: **[6 Hrs]**

Support Vector Machine, Maximal Margin Classifier, Support Vector Classifier, Kernel Trick in SVM, Naive Bayes, Naive Bayes, KNN, KNN.

Unit 4: Decision Tree: **[6 Hrs]**

Basic tree terminology, pruning trees, hyper parameter tuning, Decision trees for Regression, Decision Tree for Classification

Unit 5: Dimensionality reduction and Feature Selection: **[6 Hrs]**

Entropy, Information Gain and Gini Index, Dimensionality reduction using PCA.

Unit 6: Performance metrics for classification, Thresholding, ROC and AUC curve **[6 Hrs]**

Suggested List of Assignments:

1. Getting started with scikit learn library and understand various modules for data preprocessing, feature engineering and data modeling.
2. Implementing Linear Regression to solve a regression problem.
3. Implementing Logistic Regression to solve a classification problem.
4. Implementing Decision Trees for Regression problems.
5. Implementing Decision Trees for Classification problems.
6. Solving an End-to-End ML case study for a regression problem.
7. Solving an End-to-End ML case study for classification problems.

Books:

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning, Springer, ISBN 978-1-4614-7137-0

2. Giuseppe Bonaccorso, "Machine Learning Algorithms", Packt Publishing Limited, ISBN10: 1785889621, ISBN-13: 978-1785889622 2.

3. Tom Mitchell "Machine Learning" McGraw Hill Publication, ISBN :0070428077 9780070428072

Data Engineering

Teaching Scheme
Lectures: 3 hrs/week
End-Sem Exam – 60 marks

Examination Scheme
T1, T2 – 20 marks each

Course Outcomes:
Students will be able to:
1. Understand data processing systems, data warehousing, ETL for efficient data management.
2. Use tools like Hadoop, MapReduce with cloud platforms for data processing and storage.
3. Design secure data systems, implement access control using relevant tools.
4. Build ML pipeline for Data Visualization and Business Intelligence.

Unit 1– Introduction to Data Engineering and Data Warehousing   [6 Hrs]

Architecting Data Processing Systems / Data Pipelines, Data Infrastructure, Data Warehousing and ETL: Data modeling and database management, Data Ingestion / Extraction, Loading data, Transforming Data& Orchestration, Data Lakes, Data Reconciliation and validation

Unit 2 – Big Data and Cloud Computing   [6 Hrs]

Tools used for processing large datasets, Introduction to Hadoop and MapReduce Programming, Introduction to Apache Spark, NoSQL Databases, MongoDB, Data Ingestion with Apache Sqoop and Apache Flume, Hive & Querying, Analytics using PySpark, Major Cloud computing platforms (AWS, Google, Microsoft Azure), Cloud databases, Cloud storage, Explanation to Lab for data processing system implementation using a cloud platform

Unit 3 - Data Governance, security and privacy   [6 Hrs]

Designing for security and compliance, Access Control, Data Quality Control, Data Policies, Legal Compliance, Data Governance Tools

Unit 4- Data Validation and Best Practices in Pipelines   [6 Hrs]

Validate Early, Validate Often, A Simple Validation Framework, Validation Test Examples, Data Reconciliation, Best Practices in Pipelines: Automation, Handling Changes in Source Systems, Metadata based jobs

Unit 5: Data visualization and business intelligence   [6 Hrs]

Techniques and tools used for data representations, Overview of Machine Learning, Leveraging pre-built ML models, Building ML pipeline.

Unit 6 – Measuring and Monitoring Data Pipeline Performance   [6 Hrs]

Key Pipeline Metrics, Prepping the Data Warehouse, Logging and Ingesting Performance Data, Transforming Performance Data, Orchestrating a Performance Pipeline, Performance Transparency

Books and Information Links:
•   Data Pipelines Pocket Reference: Moving and Processing Data for Analytics by James Densmore, OREILLY
•   Professional Data Engineer Certification exam guide, Google Cloud
(https://cloud.google.com/certification/guides/data-engineer)

DS (DE-I) Foundations of Statistics and Linear Algebra

Teaching Scheme

Lectures: 3 hrs/week

End-Sem Exam – 60 marks

Examination Scheme

T1, T2 – 20 marks each

Course Outcomes:

Students will be able to:

1. Understand statistical foundations for different types of data.
2. Apply various data distributions.
3. Analyze and apply statistical tests for data inferencing.
4. Comprehend the concepts of Linear Algebra

Contents:

Unit 1: Introduction to Statistical Thinking: [6 Hrs]

Population and Sample, Parameters and statistics, variables and organization of data, Visualizing Data, Frequency Distribution, Histogram, visualizing qualitative and quantitative data.

Unit 2: Descriptive Statistics: [8 Hrs]

Measures of Central Tendency, measures of variability, Distributions, Normal Distribution, standard normal distribution, Poisson Distribution and Binomial Distribution, Sampling Distribution, Estimation, Confidence Interval.

Unit 3: Inferential Statistics: [8 Hrs]

Hypothesis Testing, Type-I and Type-II errors, Steps in hypothesis testing, one and two-tailed tests, T-Test, one sample t-test, two sample t-test, independent sample t-test, ANOVA, within and between subjects factors, one factor anova, Correlation and Regression, Chi-Square Test, contingency table.

Unit 4: Linear Algebra: [6 Hrs]

Intro to Linear Algebra, Scalar, Vectors, Matrix and Tensors, Linear Dependence and Span, Norms, SVD, Eigenvalues and Eigenvectors

Books:

1. Gilbert Strang; Introduction to Linear Algebra, Wellesley-Cambridge Press; 5th Edition; ISBN 978-0-9802327-7-6
2. Deborah J. Rumsey; Statistics for Dummies, Wiley Publishing, Inc. 2nd Edition;ISBN-13: 978-1119293521, ISBN-10: 1119293529
3. Jason Brownlee; Basics of Linear Algebra for Machine Learning; 1st Edition (ebook)

DS (DE-I) Data Visualization with Tableau

Teaching Scheme   **Examination Scheme**
Lectures: 3 hrs/week   T1, T2 – 20 marks each
                                                            End-Sem Exam – 60 marks

Course Outcomes:
Students will be able to:
1. Comprehend basic concepts in Tableau and organize data for processing.
2. Create charts and plots, Build calculations and maps using Tableau.
3. Utilize Dashboard and analytics in Tableau.

Contents:
Unit 1: Data Connections: **[6 Hrs]**
Connect to Tableau Server, Describe connection options, Connect to different data source types, from single and multiple databases, Prepare Data for Analysis: Blending Metadata Grid, Pivot, Union, Data Interpreter, Explain data extract formats and capabilities, Create extracts with multiple tables Explain performance considerations between blends, joins, and cross-database joins, Use Automatic & Custom Split.

Unit 2: Organizing & Simplifying Data: **[6 Hrs]**
Understand how to: Filter data, Sort data, Build
groups, Build hierarchies, Build sets.

Unit 3: Field & Chart Types: **[6 Hrs]**
Difference between measures and dimensions, difference between discrete and continuous fields, Tableau-generated fields, Understand how and when to
build: Histograms, Heat maps, Tree maps, Bullet graphs, Combined axis charts, Dual axis charts,Scatter plots, Cross tabs, Bar in bar charts, Box plots, Use titles, captions and tooltips effectively, Edit axes, Use mark labels and annotations.

Unit 4: Calculations: **[6 Hrs]**
Manipulate string and date calculations, Create quick table calculations,
Use level of detail (LOD) expressions, Explain different types of LOD expressions, Use Ad-hoc calculations, Work with aggregation options, Build logic statements, Build arithmetic calculations, Build grand totals and sub-totals, Use calculations in join clauses.

Unit 5: Mapping: **[6 Hrs]**
Navigate maps, including: Pan & Zoom, Filtering, Map layering, Custom
territories, Geographic search, Modify locations within Tableau, Import and manage custom geocoding, Use a background image map, Connect to spatial files.

Unit 6: Analytics: **[6 Hrs]**
Reference Lines, Reference Bands, Trend Lines, Trend Model, Forecasting,
Drag & Drop Analytics, Box Plot, Reference distributions, Statistical summary card, Instant Analytics, Data Highlighter.

Unit 7: Dashboards: **[6 Hrs]**

Build dashboards and stories, Create dashboard actions, Designdashboards for viewing on devices, Utilize visual best practices for viewing on devices, Describe publishing & sharing options.

Books:
1. Lindy Ryan, Visual Data Storytelling with Tableau, 1st Edition by Pearson, ISBN: 9789353063597
2. Milligan Joshua N., Learning Tableau, Packt Publishing Limited, ISBN: 9781784391164.
3. Ashutosh Nandeshwar , "Tableau Data Visualization Codebook", Packt Publishing, ISBN 978-1-84968-978-6

DS (DE-I) Artificial Intelligence

Teaching Scheme   **Examination Scheme**
Lectures: 3 hrs/week   T1, T2 – 20 marks each
                                                       End-Sem Exam – 60 marks

Course Outcomes:
Student will be able to:
1. Apply basic search techniques for problem solving.
2. Explain how to represent Knowledge required for problem solving.
3. Apply reasoning to sift through data.
4. Utilize AI for application in real world.

Unit I: Introduction: [6 Hrs]
Artificial Intelligence, AI Problems, AI Techniques, The Level of the Model, Criteria For Success. Defining the Problem as a State Space Search, Problem Characteristics, Production Systems, Search: Issues in The Design of Search Programs, Un-Informed Search, BFS, DFS; Heuristic Search Techniques: Generate-And- Test, Hill Climbing, Best-First Search, A* Algorithm, Problem Reduction, AO*Algorithm, Constraint Satisfaction, Means-Ends Analysis

Unit II: Knowledge Representation: [6 Hrs]
Procedural Vs Declarative Knowledge, Representations & Approaches to Knowledge Representation, Forward Vs Backward Reasoning, Matching Techniques, Partial Matching, Fuzzy Matching Algorithms and RETE Matching Algorithms

Unit III: Symbolic Logic: [6 Hrs]
Propositional Logic, First Order Predicate Logic: Representing Instance and isa Relationships, Computable Functions and Predicates, Syntax & Semantics of FOPL, Normal Forms, Unification &Resolution, Representation Using Rules, Natural Deduction; Structured Representations of Knowledge: Semantic Nets, Partitioned Semantic Nets, Frames, Conceptual Dependency, Conceptual Graphs, Scripts, CYC

Unit IV: Reasoning under Uncertainty: [6 Hrs]
Introduction to Non-Monotonic Reasoning, Truth Maintenance Systems, Logics for Non-Monotonic Reasoning, Model and Temporal Logics; Statistical Reasoning: Bayes Theorem, Certainty Factors and Rule-Based Systems, Bayesian Probabilistic Inference, Bayesian Networks, Dempster-Shafer Theory, Fuzzy Logic: Crisp Sets, Fuzzy Sets, Fuzzy Logic Control, Fuzzy Inferences & Fuzzy Systems

Unit V: Natural Language Processing: [6 Hrs]
Role of Knowledge in Language Understanding, Approaches Natural Language Understanding, steps in The Natural Language Processing, Syntactic Processing and Augmented Transition Nets, Semantic Analysis, NLP Understanding Systems; Planning: Components of a Planning System, Goal Stack Planning, Hierarchical Planning, Reactive Systems

Unit VI: Machine Learning: [6 Hrs]

Knowledge and Learning, learning by Advise, Examples, learning in problem Solving, Symbol Based Learning, Explanation Based Learning, Version Space, ID3 Decision Based Induction Algorithm, Unsupervised Learning, Reinforcement Learning, Supervised Learning: Perceptron Learning, Back propagation Learning, Competitive Learning, Hebbian Learning.

Text Books:
1. Artificial Intelligence, George F Luger, Pearson Education Publications
2. Artificial Intelligence, Elaine Rich and Knight, Mcgraw-Hill Publications

References:
1. Introduction To Artificial Intelligence & Expert Systems, Patterson, PHI 2.
2. Multi Agent systems- a modern approach to Distributed Artificial intelligence, Weiss.G, MIT Press.
3. Artificial Intelligence: A modern Approach, Russell and Norvig, Printice Hall

MLC- Research Methodology and Intellectual Property Rights

Teaching Scheme   **Evaluation Scheme:**
Self Study: 2 hrs/week   **Assignments/Presentation/Quiz/Test**

Course Outcomes (COs):
   Student will be able to
1. Understand research problem formulation and approaches of investigation of solutions for research problems
2. Learn ethical practices to be followed in research
3. Apply research methodology in case studies
4. Acquire skills required for presentation of research outcomes (report and technical paper writing, presentation etc.)
5. Infer that tomorrow's world will be ruled by ideas, concept, and creativity
6. Gather knowledge about Intellectual Property Rights which is important for students of engineering in particular as they are tomorrow's technocrats and creator of new technology
7. Discover how IPR is regarded as a source of national wealth and mark of an economic leadership in context of global market scenario
8. Study the national & International IP system
9. Summarize that it is an incentive for further research work and investment in R & D, leading to creation of new and better products and generation of economic and social benefits

Unit I: **[5 Hrs]**
Meaning of research problem, Sources of research problem, Criteria Characteristics of a good research problem, Errors in selecting a research problem, Scope and objectives of research problem. Approaches of investigation of solutions for research problem, data collection, analysis, interpretation, necessary instrumentations.

Unit II: **[5 Hrs]**
Effective literature studies approaches, analysis Use Design of Experiments /Taguchi Method to plan a set of experiments or simulations or build prototype Analyze your results and draw conclusions or Build Prototype, Test and Redesign

Unit III: **[5 Hrs]**
Plagiarism, Research ethics Effective technical writing, how to write report, Paper. Developing a Research Proposal, Format of research proposal, a presentation and assessment by a review committee

Unit IV: **[4 Hrs]**
Introduction to the concepts Property and Intellectual Property, Nature and Importance of Intellectual Property Rights, Objectives and Importance of understanding Intellectual Property Rights

Unit V: **[7 Hrs]**
Understanding the types of Intellectual Property Rights: -Patents-Indian Patent Office and its Administration, Administration of Patent System – Patenting under Indian

Patent Act , Patent Rights and its Scope, Licensing and transfer of technology, Patent information and database. Provisional and Non Provisional Patent Application and Specification, Plant Patenting, Idea Patenting, Integrated Circuits, Industrial Designs, Trademarks (Registered and unregistered trademarks), Copyrights, Traditional Knowledge, Geographical Indications, Trade Secrets, Case Studies

Unit VI: [4 Hrs]
New Developments in IPR, Process of Patenting and Development: technological research, innovation, patenting, development, International Scenario: WIPO, TRIPs, Patenting under PCT

Reference Books:

1. Aswani Kumar Bansal : Law of Trademarks in India
2. B L Wadehra : Law Relating to Patents, Trademarks, Copyright,
   a. Designs and Geographical Indications.
3. G.V.G Krishnamurthy : The Law of Trademarks, Copyright, Patents and
   a. Design.
4. Satyawrat Ponkse: The Management of Intellectual Property.
5. S K Roy Chaudhary & H K Saharay : The Law of Trademarks, Copyright, Patents
6. Intellectual Property Rights under WTO by T. Ramappa, S. Chand.
7. Manual of Patent Office Practice and Procedure
8. WIPO : WIPO Guide To Using Patent Information
9. Resisting Intellectual Property by Halbert ,Taylor & Francis
10. Industrial Design by Mayall, Mc Graw Hill
11. Product Design by Niebel, Mc Graw Hill
12. Introduction to Design by Asimov, Prentice Hall
13. Intellectual Property in New Technological Age by Robert P. Merges, Peter S. Menell, Mark A. Lemley

MLC-Effective Technical Communication

Teaching Scheme: **Evaluation Scheme:**
Self Study: 1hr / week   **Assignments/Presentation/Quiz/Test**

Course Outcomes (COs):
Student will be able to
1. Produce effective dialogue for business related situations
2. Use listening, speaking, reading and writing skills for communication purposes and attempt tasks by using functional grammar and vocabulary effectively
3. Analyze critically different concepts / principles of communication skills
4. Demonstrate productive skills and have a knack for structured conversations
5. Appreciate, analyze, evaluate business reports and research papers

Unit I: Fundamentals of Communication   [4 Hrs]
7 Cs of communication, common errors in English, enriching vocabulary, styles and registers

Unit II: Aural-Oral Communication   [4 Hrs]
The art of listening, stress and intonation, group discussion, oral presentation skills

Unit III: Reading and Writing   [4 Hrs]
Types of reading, effective writing, business correspondence, interpretation of technical reports and research papers

Reference Books

1. Raman Sharma, "Technical Communication", Oxford University Press.
2. Raymond Murphy "Essential English Grammar" (Elementary & Intermediate) Cambridge University Press.
3. Mark Hancock "English Pronunciation in Use" Cambridge University Press.
4. Shirley Taylor, "Model Business Letters, Emails and Other Business Documents" (seventh edition), Prentise Hall
5. Thomas Huckin, Leslie Olsen "Technical writing and Professional Communications for Non- native speakers of English", McGraw Hill.

Reference books/paper(s):
1. D.J.C. MacKay, "Information Theory, Inference, and Learning Algorithms", Cambridge University Press
2. C. E. Shannon, A Mathematical Theory of Communication, Bell Sys. Tech Journ, 1948. (available online)

Web Resources:
1. NPTEL Course (Information Theory and Coding – IIT, Bombay) : http://nptel.ac.in/syllabus/117101053/
2. MIT OpenCourseWare (Information Theory) : http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-441- information-theory-spring-2010/index.htm

[OE] Data Structures

Teaching Scheme:

Lectures: 3 Hrs/week

Examination Scheme:

T1 and T2: 20 Marks each

End-Sem Exam: 60 Marks

Course Outcomes:

Students will be able to:

1. Apply and implement advanced data structures, such as B-trees, multi-way trees, balanced trees, heaps, priority queues, to solve computational problems
2. Analyze the time and space complexity of advanced data structures and their supported operations
3. Compare the time and space tradeoff of different advanced data structures and their common operations

Unit I: **[6 Hrs]**

Review of Basic Concepts: Abstract data types, Data structures, Algorithms, Big Oh, Small Oh, Omega and Theta notations, Solving recurrence equations, Master theorems, Generating function techniques, Constructive induction.

Unit II: **[8 Hrs]**

Advanced Search Structures for Dictionary ADT: Splay trees, Amortized analysis, 2-3 trees, 2-3-4 trees, Red-black trees, Randomized structures, Skip lists, Treaps, Universal hash functions.

Unit III: **[6 Hrs]**

Advanced Structures for Priority Queues and Their Extensions: Binary Heap, Min Heap, Max Heap, Binomial heaps, Leftist heaps, Skewed heaps, Fibonacci heaps and its amortized analysis, Applications to minimum spanning tree algorithms.

Unit IV: **[6 Hrs]**

Data Structures for Partition ADT: Weighted union and path compression, Applications to finite state automata minimization, Code optimization.

Unit V: **[6 Hrs]**

Graph Algorithms: DFS, BFS, Biconnected components, Cut vertices, Matching, Network flow; Maximum-Flow / Minimum-Cut; Ford–Fulkerson algorithm, Augmenting Path

Unit VI: **[8 Hrs]**

Computational Geometry: Geometric data structures, Plane sweep paradigm, Concurrency, Java Threads, Critical Section Problem, Race Conditions, Re-entrant code, Synchronization; Multiple Readers/Writers Problem

Text Books:

1. Introduction to Algorithms; 3rd Edition; by by Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein; Published by PHI Learning Pvt. Ltd. ; ISBN-13: 978-0262033848 ISBN-10: 0262033844
2. Algorithms; 4th Edition; by Robert Sedgewick and Kevin Wayne; Pearson Education, ISBN-13: 978-0321573513

References:

1. Algorithms; by S. Dasgupta, C.H. Papadimitriou, and U. V. Vazirani; Published by Mcgraw-Hill, 2006; ISBN-13: 978-0073523408 ISBN-10: 0073523402
2. Algorithm Design; by J. Kleinberg and E. Tardos; Published by Addison-Wesley, 2006; ISBN-13: 978-0321295354 ISBN-10: 0321295358

Big Data Analytics with Apache Spark ()

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
Lab: 2hrs/week   End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Set up Apache Spark, Hadoop ecosystems.
2. Use Spark SQL to dataframes for Big Data
3. Build models using KAFKA workflow.
4. Use Spark with ML

Contents:
Unit 1: Introduction to Spark: **[6 Hrs]**
Introduction to Apache Hadoop, Overview of Hadoop Ecosystem, Spark – Introduction, Spark –
Ecosystem Components. Spark basics: Features and Use Cases, SparkContext, Stage, Executor.

Unit 2: Working with RDDS in Spark: **[8 Hrs]**
Spark – RDD, Spark – Ways to Create RDD, Spark – RDD Persistence & Caching, Spark –
RDD Features, Spark – Paired RDD, Spark – RDD limitations, Spark – Transformations
Actions, Spark – RDD Lineage.

Unit 3: Spark SQL AND dataframe: **[8 Hrs]**
Spark SQL – Introduction, Spark SQL – Features, Spark SQL – DataFrame, Spark SQL –
DataSet, Spark SQL – Optimization, HIVE Fundamentals.

Unit 4: Spark Configuration, Monitoring and Tuning: **[6 Hrs]**
Spark – In-Memory Computation, Spark – Directed Acyclic Graph, Spark – Cluster Managers,
Spark – Performance Tuning, RDD vs DataFrame vs DataSet.

Unit 5: KAFKA: **[6 Hrs]**
Kafka – Introduction, Kafka Architecture, Kafka Workflow, Kafka – Cluster configuration,
Kafka monitoring tools.

Unit 6: Spark Streaming and ML: **[8 Hrs]**
 Spark Streaming – Introduction, Spark Streaming – DStream, Spark Streaming –
Transformations, Spark Streaming – Checkpointing, Spark – Batch vs Real Time, Spark MLib.

Books:
1. Rudy Lai, Bartłomiej Potaczek, Hands-On Big Data Analytics with PySpark, Packt
Publishing Limited, ISBN:978-1-83864-413-0.
2. Jeffrey Aven, Data Analytics with Spark Using Python, Pearson Publications, 1st Edition,
ISBN-13: 978-0134846019

Suggested List of Assignments:

1. Set up a Hadoop cluster using HDFS and execute basic file operations (upload, delete, list). Run a MapReduce job on Hadoop to analyze a sample dataset.
2. Install and configure Apache Spark in a standalone mode.
3. Understand Resilient Distributed Datasets (RDDs) in Spark and perform transformations and actions (map, filter, reduce, collect).
4. Use Spark SQL to query and analyze structured data stored in various formats (CSV, JSON, Parquet).
5. Optimize Spark SQL queries using techniques such as partitioning, bucketing, and caching.
6. Set up a Kafka cluster and configure topic partitions and replication factors. Produce and consume messages using Kafka producers and consumers.
7. Monitor Kafka cluster metrics using tools like Kafka Manager or Confluent Control Center.
8. Implement a Spark Streaming application using DStreams to process real-time data from a Kafka topic.
9. Apply transformations (map, flatMap, window) on DStreams for data processing.

Advanced Machine Learning and Deep Learning Fundamentals ()

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
Lab: 2hrs/week   End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Apply clustering algorithms for dense as well as sparse datasets.
2. Apply association rule mining for frqeuent itenmset generation and Utilize ensemble methods for efficient machine learning.
3. Comprehend Deep Learning concepts and models.
4. Apply CNN and RNN neural networks for machine learning.

Contents:

Unit 1: Unsupervised Learning: **[6 Hrs]**
Content based, collaborative and hybrid filtering techniques, Clustering, Hierarchical Clustering, distance metrics, inter-cluster and intra-cluster distances, dendogram, kmeans clustering, k means++ clustering, elbow method for determining number of clusters.

Unit 2: Feature Reduction and Recommender systems: **[8 Hrs]**
Principal Component Analysis (PCA), First Principal Component, Eigenvalues and PCA, Association Rule Mining, Association Rule Mining: Market Basket Analysis, Association Rule Generation: Apriori Algorithm, Apriori Algorithm Example: Part A, Apriori Algorithm Example: Part B, Apriori Algorithm: Rule Selection

Unit 3: Regularization and Ensemble Methods: **[8 Hrs]**
Subset Selection Methods, Regularization, ridge regularization and lasso regularization, variable selection lasso regularization. Applying cross validation for learning rate, Bias-Variance Tradeoff, Cross Validation and Boot Straping, Ensemble Methods (Bagging), Ensemble Methods (Random Forest), Ensemble Methods (Boosting)

Unit 4 : Introduction to Deep Learning: **[6 Hrs]**
What is Deep Learning?, Multilayer Perceptron ,Feed forward neural, Back propagation, Gradient descent, Vanishing gradient problem, *Activation Functions:* RELU, LRELU, ERELU, Optimization Algorithms, *Hyper parameters*: Layer size, Magnitude (momentum, learning rate),Regularization (dropout, drop connect, L1, L2)

Unit 5: Convolutional Neural Network: **[8 Hrs]**
Introduction to CNN, Convolution Operation, Parameter Sharing, Equivariant Representation, Pooling, Variants of the Basic Convolution Function, The basic Architecture of CNN, Popular CNN Architecture – AlexNet.

Unit 6: Applications of Deep Learning :   **[6 Hrs]**

Overview of Deep Learning Applications: Image Classification, Social N/w/ analysis, Speech Recognition, Recommender system, Natural Language Processing.

Suggested List of Assignments:
1. Implementing Clustering Algorithms for Machine Learning.
2. Build an ensemble model to correctly classify the outcome variable by demonstrating the use of Bagging, Boosting, Stacking.
3. Applying cross validation on a machine learning algorithm.
4. Model a multi-layer neural network model to demonstrate deep learning application.

Books:
1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning, Springer, ISBN 978-1-4614-7137-0
2. Giuseppe Bonaccorso, "Machine Learning Algorithms", Packt Publishing Limited, ISBN10: 1785889621, ISBN-13: 978-1785889622 2.
3. Tom Mitchell "Machine Learning" McGraw Hill Publication, ISBN :0070428077 9780070428072.
4. Jack Zurada: Introduction to Artificial Neural Systems, PWS Publishing Co. Boston, 2002
5. Josh Patterson, Adam Gibson, "Deep Learning: A Practitioners Approach", O"REILLY, SPD, ISBN: 978-93-5213-604-9, 2017 Edition 1st .
6. François Chollet, Deep Learning with Python, Manning Publications, ISBN: 9781617294433
7. Antonio Gulli,Sujit Pal, Deep Learning with Keras, Packt Publishing Limited, ISBN: 978-1-78712-842-2

MLOps and Systems ()

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
Lab: 2hrs/week   End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Build and validate well known ML/DL model prototypes on a variety of ML use cases such as Data Streams
2. Apply transfer learning for ML Model deployment on cloud platforms.
3. Understand the impact of data drift and concept drift in ML pipelines.
4. Apply AutoML and collaborative frameworks such as MLflow.
5. Build and maintain CI/CD pipelines for cloud-based ML-Model Deployments.
6. Apply production-specific software services using Prometheus, grafana cloud, elk stack (AWS), fluentD and cloudwatch (AWS)

Unit I: Statistical ML Models for Data Streams   [8 Hrs]
Classification, Regression, Unsupervised learning, AutoML: autosklearn, TPOT, Linear and non-Linear Models: linear regression, Random Forest, SVM, kNN, k-means, logistic regression, Visualization using D3, Tableau.

Unit II: ML API Development and Deployment on GCP and AWS   [6 Hrs]
Flask, FastApi, Tensorflow serving, Tensorflow lite for optimization latency

Unit III: Cloud-based APIs   [6 Hrs]
Intro to Kubeflow, Collaborative platforms: MLflow, AutoML in GCP, Azure, AWS, Hyper parameter optimization

Unit IV: ML Deployment and MLOps Pipelines   [8 Hrs]
ML Deployment : Docker Containers, Kubernetes, FluentD, Elk Stack
MLOps Pipelines : CI/CD Pipelines, Gitlab, Jenkins, YAML/XML, Profiler

Unit V: Application Production Testing on Google Optimize   [6 Hrs]
Canary pattern, A/B Pattern, Shadow Pattern

Unit VI: Production Software Build and Analytics   [8 Hrs]
Prometheus, Grafana Cloud, Cloudwatch, Special Topics : Building models from scratch, Hyper-parameterization, optimization, quantization

Suggested Assignments

1**.** E-commerce use-case for data wrangling, compression, meta-data tagging, metrics and visualization
2. ML API Development and Deployment on GCP and AWS : Rapid Prototyping for NLP use case
3. Computer Vision use case: Hyper-parameterization using MLflow
4. Study of ML Deployment use case
5. Study of MLOps Pipelines Deployment use case
6. Application Production Testing on Google Optimize Deployment use case
7. Production Software Build and Analytics Deployment use case
Books

1. Mark Treveil, CL Stenac, L Dreyfus-Schmidt, Kenji LeFevre, Nicolas Omont, Introducing MLOps: How to Scale Machine Learning in the Enterprise, O'REILLY, 2021
2. Noah Gift and Alfredo Deza, Practical MLOps: Operationalizing Machine Learning Models, O'REILLY, 2021.
3. Hannes Hapke and Catherine Nelson, Building Machine Learning Pipelines: Automating Model Life Cycles with TensorFlow, O'REILLY, 2020

DS (DE1) Time Series Data Analysis

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
                                                             End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Understand the inherit difference with normal data and time series data.
2. Perform various analysis on time series data.
3. Derive conclusion from time series data.

Contents:
Unit :1 Introduction to Time Series: [8 Hrs]
Trend Effect, Seasonal effect, Cyclic effect, Components of Time series, Additive Model, Naive/Snaive, Multiplicative Model, Autocorrelation function and white noise, Sample ACVF, Sample ACF, PACF, Correlogram.

Unit 2: Basic Forecasting Models: [8 Hrs]
Naïve Approach, Moving Average Method, Simple exponential Smoothing method, Holt's Linear Trend Model, Holt's Winter Model.

Unit 3: Single and Multivariate Time Series: [8 Hrs]
Stationary, Differencing, Auto-regressive (AR), Moving Average (MA), Auto-regressive moving average (ARMA), Auto-regressive integrated moving average (ARIMA), Seasonal ARIMA (SARIMA), Granger's Causality Test, Vector Autoregression (VAR), Augmented Dickey-fuller test (ADF), Random Walk,YuleWalker estimation, FPE, AIC, AICc, BIC, residual analysis and diagnostic checking.

Unit 4: Time Series using Deep Learning: [8 Hrs]
 Prepare Time Series Data for CNNs and LSTMs, MLPs for Time Series Forecasting, CNNs for Time Series Forecasting, Simple LSTM, Stacked RNN and Bidirectional LSTMs, RNN.

Books:

1.Introduction to Time Series and Forecasting, 2nd Edition; Peter J. Brockwell, Richard A. Davis; ISBN 0-387-95351-5; Springer-Verlag New York, Inc.
2.Practical Time Series Analysis, 1st Edition; Aileen Nielsen; ISBN: 978-1-492-04165-8; O'Reilly Media, Inc.
3.Rob J Hyndman and George Athanasopoulos, Forecasting: Principles and Practice, OTexts, ISBN-0987507109.

DS (DE1) Computer Vision

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
                                                                             End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Perform Digital Image processing techniques and Extract images from features.
2. Apply Image segmentation and Transformation techniques.
3. Perform Pattern and Motion Analysis by using Transformers.

Contents:

Digital Image Processing: **[6 Hrs]**
Fundamentals, Types of Image Processing, Image Acquisition Methods, Human Perception of Color and Images , Phases of computer Vision and its Applications, Different color Models: RGB ,YCbCr , CMYK ,HSV ,LAB , YIQ and their conversion, Relation Between a pixels: Neighborhood, Adjacency/connectivity, Connected components, Region and Boundaries, Arithmetic logic operations on pixels, Distance Measures.

Low-level Image Processing: **[6 Hrs]**
Image Enhancement in Spatial Domain -Histogram Processing, Contrast Stretching, Log Transformation, Gamma Correction, Smoothing and Sharpening; Logical and Arithmetic Operations, Morphological Image Processing, Image Enhancement in Frequency Domain, Fourier Transform, Convolution and Filtering, Image Restoration.

Image Feature Extraction: **[6 Hrs]**
Edge detection – Canny, Sobel, Prewitt, LOG, DOG, Line detector: Hough Transform; Corner detectors – Harris and Hessian Affine; Orientation Histogram, SIFT, SURF, HOG, GLOH, Gaussian derivative filters, Gabor Filters and DWT, Histogram processing, Histogram Equalization Histogram Specification

Image Segmentation: **[6 Hrs]**
Edge Based Approaches to Segmentation, Texture Segmentation, Object Detection and Segmentation , Similarity based: Thresholding, Region growing , Region splitting and merging, Discontinuity-based: Use of point, line and edge for segmentation.

Pattern Analysis: **[6 Hrs]**
Clustering: K-Means; Gaussian Mixture Model (GMM); Classification – Discriminant Function, Supervised, Semi-supervised, Unsupervised; Classifiers: Bayes, KNN, ANN models; Dimensionality Reduction: PCA, LDA, ICA; Non-parametric methods, Vision with Deep Learning Approaches: CNN and Transformers

Motion Analysis: **[6 Hrs]**
Background Subtraction and Modeling, Optical Flow, KLT, Spatio-Temporal Analysis.

Applications and Performance Measures: [6 Hrs]

Image classification, Recognition, Biometrics, Augmented Reality, Security and Surveillance, Performance Evaluation Measures.

Books:

1. Computer Vision: A Modern Approach, D. A. Forsyth and J. Ponce, Pearson Education, 2003. (693 pages), ISBN: 9780130851987.
2. Computer Vision: Algorithms and Applications, Richard Szeliski, Springer-Verlag, 2011. (832 pages), ISBN: 978-1848829343.
3. Digital Image Processing, Rafael C. Gonzalez and Richard E. Woods, Pearson Education, 2008. (976 pages), ISBN: 9788131726952.

DS (DE1) R Programming

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
                                                                              End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Perform vector operations and matrix operations using R programming.
2. Apply different R programming structures like conditions and control functions.
3. Manipulate data in files using R programming

Contents:
Unit 1: Introduction to R Programming: [6 Hrs]
Getting started with RStudio, workspace in R, packages in R, installing packages in R, Built-in Datasets in R, defining variables, constants and strings, operators in R

Unit 2: Data Structures and Functions: [6 Hrs]
Vectors, operations on vectors, generating vectors using seq(), Repeating Vector Constants with rep(), NA and NULL values, arrays and matrices, matrix indexing, matrix operations, List, list functions, data frames, extracting sub-data frames, using the rbind() and cbind() functions and alternatives, factors, common functions used with factors, split() and by()

Unit 3: R Programming Structures: [6 Hrs]
Conditions and loops, control functions, apply family, default values for arguments, lazy evaluation in functions, objects and classes, debugging

Unit 4: Data Manipulation in R: [6 Hrs]
Importing and exporting data in R, Reading text files, writing and saving data objects to file in R, data Manipulation using dplyr package, exploratory data analysis using R, data cleaning, data cleaning, ggplot2 package, statistics with R, machine learning implementation in R

Books:
1. Norman Matloff, "The art of R programming: A tour of statistical software design", No Starch Press, ISBN: 9781593273842
2. Andrie De Vries and Joris Meys, "R for Dummies", John Wiley & Sons, ISBN: 9781119962847
3. Sandip Rakshit, "R Programming for Beginners", McGraw Hill India, Edition 1, 2017, ISBN: 9789352604555
4. Hadley Wickham, Garrett Grolemund, "R for data science : Import, Tidy, Transform, Visualize, And Model Data", O'Reilly, ISBN: 9781491910399
5. Andy Field, "Discovering Statistics Using R", Sage publications, ISBN: 9781446200469

DS (DE-II) Generative Adversarial Networks

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
                                                     End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Understand various clocks of Generative Adversarial Networks (GANs)
2. Analyse and understand Deep convolution architecture with its challenges pertaining to GANs.
3. Design various structure of GANs while optimizing the various training challenges.
4. Study, implement and understand various application of GANs for image, text, and speech datasets.

Contents:
Unit I: Introduction to GANs :   **[10 Hrs]**
Generative Adversarial Networks (GANs), mathematics underpinning, Random noise vector, Generator network, Discriminator network, Iterative training/tuning, Adversarial training, Nash equilibrium, Introduction to generative modelling, Encoder network, Latent space, Decoder network, Generation using an autoencoder, Variational autoencoder

Unit II: Deep Convolutional GAN: **[10 Hrs]**
Convolutional neural networks, Convolutional filters, Parameter sharing, Batch normalization, Training challenges (Mode collapse, Slow convergence, Overgeneralization), Non-Saturating GAN

Unit III: Various types of GANs: **[15 Hrs]**
Semi-Supervised GAN: Introduction to Semi-Supervised GAN, Architecture, Training process, Training objective
Conditional GAN: Motivation, CGAN Generator, CGAN Discriminator, Architecture of CGAN
CycleGAN: Image-to-image translation, Cycle-consistency loss, Adversarial loss, Identity loss, CycleGAN architecture: building the network Object-oriented design of GANs

Unit IV: Real-World Applications for GANs: **[15 Hrs]**
Human Faces Generation, Deep Fake, Image-to-Image Translation/Restoration, Text to Image Generation, Enhancing Image Resolution, Semantic Image Inpainting, Text to Speech, Speech **Enhancement** with GANs

Reference Books

1. Foster, D., 2019. Generative Deep Learning. Teaching Machines to Paint, Write, Compose and Play (2019). Beijing-Boston-Farnham-Sebastopol-Tokyo, OREILLY, p.330.
2. Bok, Vladimir, and Jakub Langr. GANs in action: deep learning with generative adversarial networks. Simon and Schuster, 2019.
3. Ganguly, Kuntal. Learning generative adversarial networks: next-generation deep learning simplified. Packt Publishing, 2017.
4. Valle, Rafael. Hands-On Generative Adversarial Networks with Keras: Your guide to implementing next-generation generative adversarial networks. Packt Publishing Ltd, 2019.

DS (DE-II) Reinforcement Learning

Teaching Scheme:
Lectures: 3 Hrs/week

Examination Scheme:
T1 and T2: 20 Marks each
End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Estimate action value functions of k-armed bandit problems aimed at multiple optimalities.
2. Define RL problems, value functions and policies based on finite MDPs.
3. Utilise dynamic programming algorithms for estimating optimal value functions and policies.
4. Demostrate Monte Carlo on-ploicy and off-policy techniques for prediction and control Prove efficacy of temporal difference prediciton and control methods over earlier techniques.

Module 1: Introduction to Reinforcement Learning   (04 hrs.)
Examples, Elements of Reinforcement Learning, Limitations and Scope, An Extended Example: Tic-Tac-Toe, Early History of Reinforcement Learning

Module 2: Multi-armed Bandits   (08 hrs.)
k-armed Bandit Problem, Action-value Methods, Bandit optimalities, The 10-armed Testbed, Incremental Implementation   Tracking a Nonstationary Problem, Optimistic Initial Values, Upper-Confidence-Bound Action Selection, Gradient Bandit Algorithms, Contextual Bandits

Module 3: Finite Markov Decision Processes   (06 hrs.)
The Agent–Environment Interface, Goals and Rewards, Returns and Episodes, Continuing Tasks, Policies and Value Functions, Optimal Policies and Optimal Value Functions, Optimality and Approximation

Module 4: Dynamic Programming (DP)   (06 hrs.)
Policy Evaluation (Prediction), Policy Improvement, Policy Iteration, Value Iteration, Asynchronous DP, Generalized Policy Iteration Efficiency of DP

Module 5: Monte Carlo (MC) Methods   (06 hrs.)
MC Prediction, MC Estimation of Action Values, Monte Carlo Control, Monte Carlo Control without Exploring Starts, Off-Policy Prediction via Importance Sampling. Incremental Implementation, Off-policy Monte Carlo Control

Module 6: Temporal-Difference (TD) Learning   (06 hrs.)
TD Prediction, Advantages of TD Prediction Methods, Optimality of TD(0), Sarsa: On-policy TD Control, Q-learning, Expected Sarsa, Maximization Bias and Double Learning, Games, After-states, and Other Special Cases

Books:
1. R. S. Sutton and A. G. Barto. Reinforcement Learning - An Introduction, MIT Press, First Edition 1998 and Second Edition 2018
2. Stefano V. Albrecht, Filippos Christianos, Lukas Schäfer - Multi-agent RL, MIT Press, 2024
3. Tor Lattimore and Csaba Szepesvari - Bandit algorithms - Cambridge University Press,Online edition,

DS (DE-II) Natural Language Processing

Teaching Scheme: **Examination Scheme:**
Lectures: 3 Hrs/week   T1 and T2: 20 Marks each
                                                        End-Sem Exam: 60 Marks

Course Outcomes:
Students will be able to:
1. Demonstrate the understanding of basic text processing techniques in NLP.
2. Design, implement and evaluate part-of-speech taggers and parsers for a language.
3. Build language models and demonstrate Word Sense Disambiguation using WordNet.
4. Analyze and build word embeddings for different languages.

Unit I: Introduction [6 Hrs]
What is NLP, Fundamental and Scientific goals, Engineering goals, stages of NLP, problems in NLP, Applications of NLP, Empirical Laws of language, zipf's law, Heap's law.

Unit II: Basic Text Processing [8 Hrs]
Tokenization, word token, word type, sentence segmentation, feature extraction, issues intokenization for different languages, word segmentation, text segmentation, normalization, case folding, Spelling Correction, Morphology, Stemming, Porters Algorithm, lemmatization,spelling correction - dynamic programming approach for finding edit distance, N-gram Language Modeling- context sensitive spelling correction, probabilistic language model, auto completion prediction, Evaluation and perplexity, Smoothing techniques.

Unit III: POS Tagging [8 Hrs]
Sequence labeling tasks of NLP, POS tagging, POS tag sets, Hidden Markov ModelIntroduction, Markov Processes, HMM characterization -Likelihood of a sequence (Forward Procedure, Backward Procedure), Best state sequence (Viterbi Algorithm), Re-estimation(BaumWelch - Forward-Backward Algorithm) , Models for Sequential tagging – Maximum Entropy, Conditional Random Field.

Unit IV: Syntax [10 Hrs]
Constituency and dependency parsing, Constituency parser -Syntactic structure, Parsing methodology, Different parsing algorithms, Parsing in case of ambiguity, Probabilistic parsing, CKY algorithm, Issues in parsing, Dependency parsing- Syntactic structure, Parsing methodology, Transition-Based Dependency Parsing, Graph-Based dependency parsing, Evaluation, Co-reference resolution, Named-entity recognition.

Unit V: Knowledge Base and Semantics [6 Hrs]
WordNet: Word Senses, Word relations, Word similarity and thesaurus methods, Word sense disambiguation, WordNet. Lexical and Distributional Semantics - Introduction, models of semantics, applications.

Unit VI: Word Embeddings [6 Hrs]
Introduction, one-hot vectors, methods of generating word embeddings, Skip-gram, CBOW, Glove model, Fast Text model, evaluation measures-rough scores.

Text Books:
1. Daniel Jurafsky and James H. Martin, "Speech and Language Processing", Second Edition,

Prentice Hall, 2008, ISBN: 978-0131873216.
2. Allen James, "Natural Language Understanding", Second Edition, Benjamin/Cumming, 1994, ISBN: 978-0805303346.
3. Chris Manning and Hinrich Schuetze, "Foundations of Statistical Natural Language Processing", MIT Press, ISBN: 978-0262133609.

Reference Books:
1. Journals: Computational Linguistics, Natural Language Engineering, Machine Learning, Machine Translation, Artificial Intelligence.
2. Conferences: Annual Meeting of the Association of Computational Linguistics (ACL), Computational Linguistics (COLING), European ACL (EACL), Empirical Methods in NLP (EMNLP), Annual Meeting of the Special Interest Group in Information Retrieval (SIGIR), Human Language Technology (HLT).