COEP Technological University Pune

(A ary Public University of Govt. of Maharashtra)

NEP 2020 Compliant

Proposed Curriculum Structure

M. Tech. Data Science

(Effective from: A.Y. 2025-26)

List of Abbreviations

Abbreviation	Title	No of courses	Credits	% of Credits
PSMC	Program Specific Mathematics Course	1	4	5.00%
PSBC	Programme Specific Bridge Course	1	3	3.75%
PCC + LC	Programme Core Course + Laboratory Course	6	24	30.00%
PEC	Programme Elective Course	3	9	11.25%
OJT	On Job Training	1	3	3.75%
OE	Open Elective	1	3	3.75%
LLC	Liberal Learning Course	1	1	1.25%
SLC	Self-Learning Course	2	6	7.50%
RM	Research Methodology	1	3	3.75%
AEC	Ability Enhancement Course	1	2	2.50%
VSEC	Project	2	22	27.5
	Total	20	80	100%

Curriculum Structure for M. Tech. (Data Science)

NEP Effective from 2025-26

Semester I

Sr. No.	Course Type	Course Code	Course Name	L	T	P	S	Cr	Evaluation Scheme (Weightages in %)				
									Theory			Laboratory	
									MSE	TA	ESE	CIE	ESE
1.	PSMC	<tbd>	Probability Statistics and Queuing Theory	3	1	-	1	4	30	20	50	-	-
2.	PSBC	<tbd>	Algorithms and Complexity Theory	2	-	2	1	3	30	20	50	50	50
3.	PCC	<tbd>	SQL and Python Programming	3	-	2	1	4	30	20	50	50	50
4.	PCC	<tbd>	Machine Learning	3	-	2	1	4	30	20	50	50	50
5.	PCC	<tbd>	Data Engineering	3	-	2	1	4	30	20	50	50	50
6.	PEC-1	<tbd>	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	-	-	1	3	30	20	50	-	-
7.	RM	<tbd>	Research Methodology and Intellectual Property Rights	3	-	-	1	3	30	20	50	-	-
			Total Credits					25					

Legends: L-Lecture, T-Tutorial, P-Practical, S-Self Study, Cr-Credits, ISE: In-Semester-Evaluation, ESE: End-Semester-Evaluation, MSE: Mid-Semester Evaluation, TA: Teacher's Assessment, CIE: Continuous-Internal-Evaluation

Semester II

Sr. No.	Course Type	Course Code	Course Name	L	T	P	S	Cr	Evaluation Scheme (Weightages in %)				
									Theory			Laboratory	
									MSE	TA	ESE	CIE	ESE
1.	OE	<tbd>	Open Elective: Data Structures	3	-	-	1	3	30	20	50	-	-
2.	PCC	<tbd>	Big Data Analytics with Apache Spark	3	-	2	1	4	30	20	50	50	50
3.	PCC	<tbd>	Advanced Machine Learning and Deep Learning	3	-	2	1	4	30	20	50	50	50
4.	PCC	<tbd>	ML Ops and Systems	3	-	2	1	4	30	20	50	50	50
5.	PEC-2	<tbd>	Program Specific Elective –II 1. Time Series Data Analysis 2. Computer Vision 3. R Programming	3	-	-	1	3	30	20	50	-	-
6.	PEC-3	<tbd>	Program Specific Elective –III 1. Generative Adversarial Networks 2. Reinforcement Learning 3. Natural Language Processing 4. Industry offered Electives	3	-	-	1	3	30	20	50	-	-
7.	AEC	<tbd>	Technical Communication Skills	1	-	2	1	2	50	50	-	100	
8.	LLC	<tbd>	Liberal Learning Course	-	-	2	2	1	-	-	-	100	-
			Total Credits					**24**					

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	L	T	P	S	Cr	Evaluation Scheme (Weightages in %)				
									Theory			Laboratory	
									MSE	TA	ESE	CIE	ESE
1	SLC	<tbd>	Massive Open Online Course –I	3	-	-	1	3	-	-	100	-	-
2	SLC	<tbd>	Massive Open Online Course –II	3	-	-	1	3	-	-	100	-	-
3	OJT	<tbd>	Internship	-	-	-	-	3	-	-	100	-	-
4	Project	<tbd>	Dissertation Phase – I	-	-	22	12	11	-	-	-	70	30
			Total Credits					20					

Semester IV

Sr. No.	Course Type	Course Code	Course Name	L	T	P	S	Cr	Evaluation Scheme (Weightages in %)				
									Theory			Laboratory	
									MSE	TA	ESE	CIE	ESE
1	Project	<tbd>	Dissertation Phase – II	-	-	22	12	11	-	-	-	70	30
			Total Credits					11					

(PSMC) Probability Statistics and Queuing Theory

Teaching Scheme

Lectures: 3 Hrs/ Week

Tutorial: 1 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

1. Solve problems related to basic probability theory
2. Solve problems related to basic concepts and commonly used techniques of statistics
3. Model a given scenario using continuous and discrete distributions appropriately and estimate the required probability of a set of events
4. Apply the theory of probability and statistics to solve problems in domains such as machine learning, data mining, computer networks etc.

Course content

Basic Probability Theory **[2 Hrs]**

Probability axioms, conditional probability, independence of events, Bayes' rule, Bernoulli trials

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

Stochastic Processes **[8 Hrs]**

Introduction and classification of stochastic processes, Bernoulli process, Poisson process, Renewal processes

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

Statistical Inference **[8 Hrs]**

Parameter Estimation – sampling from normal distribution, exponential distribution, estimation related to Markov chains, Hypothesis testing.

Regression and Analysis of Variance **[6 Hrs]**

Least square curve fitting, Linear and non-linear regression, Analysis of variance.

Textbooks

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

Reference Books

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

(PSBC) Algorithms and Complexity Theory

Teaching Scheme

Lectures: 2 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

1. Understand the fundamental concepts of algorithm design and time complexity analysis.
2. Apply suitable algorithm design techniques such as divide and conquer, greedy, and dynamic programming to solve well-known problems.
3. Analyze the time and space complexity of algorithms using asymptotic notation.
4. Evaluate computational problems by classifying them into complexity classes and justifying the classification.
5. Create efficient algorithms for given computational tasks using appropriate design strategies

Course content

Mathematical Foundation **[4 Hrs]**

Growth of functions – Asymptotic notation, Standard notation and common functions, Summations, solving recurrences.

Analysis of Algorithms **[4 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.

Standard Design Techniques-I **[6 Hrs]**

Divide and Conquer, Greedy method.

Standard Design Techniques-II **[6 Hrs]**

Dynamic programming, Network flow

Standard Design Techniques-III **[6 Hrs]**

Backtracking, Branch-and-bound.

Complexity Theory **[4 Hrs]**

Introduction to NP-Completeness, Reducibility (SAT, 3VC, Independent Set, Subset Sum, Hamiltonian Circuit etc.)

Self-study **[12 Hrs]**

Sorting in linear time, Elementary graph algorithms, Minimum spanning tree, Number -Theoretic algorithms: GCD algorithm, Chinese remainder theorem, Primality testing, String Matching Algorithms

Textbooks

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

Laboratory assignments

1. **Recurrence Relations**: Study and solve recurrence relations using formal methods such as substitution, recursion tree, and the Master Theorem to determine algorithmic time complexities.
2. **Sorting Algorithm Analysis:** Implement, trace, and analyze the performance of advanced sorting algorithms including Heap Sort, Quick Sort, and Merge Sort, with emphasis on time complexity in best, worst, and average cases.
3. **Greedy Strategy Implementation**: Design and implement greedy algorithms to solve problems such as Fractional Knapsack, Job Sequencing with Deadlines, Huffman Coding, and Optimal Merge Pattern. Analyze correctness and efficiency.
4. **Graph Algorithms**: Develop and evaluate graph-based solutions including Single-source shortest path using Dijkstra’s algorithm, Minimum Spanning Tree construction using Prim’s and Kruskal’s algorithms.
5. **Dynamic Programming Techniques:** Implement dynamic programming solutions for classical problems: Matrix Chain Multiplication, Longest Common Subsequence (LCS), 0/1 Knapsack, All-Pairs Shortest Paths using Floyd-Warshall algorithm, Bellman-Ford algorithm for single-source shortest paths
6. **String Matching Algorithms:** Implement and analyze the performance of pattern matching algorithms including the Naive approach and Knuth-Morris-Pratt (KMP) algorithm.
7. **Branch and Bound Applications:** Apply the branch-and-bound strategy to solve computationally hard problems such as the 0/1 Knapsack and Travelling Salesperson Problem (TSP), focusing on pruning and bounding techniques.
8. **Backtracking Approaches:** Solve combinatorial problems using backtracking techniques, including N-Queens Problem, Graph Coloring, Hamiltonian Path, Travelling Salesperson Problem (exact approach)
9. **Network Flow Algorithms:** Implement and analyze network flow algorithms such as: Ford-Fulkerson method, PushRelabel algorithm and evaluate their time complexity and applicability.
10. **NP-Completeness and Reductions:** Apply polynomial-time reduction techniques to prove the NP-completeness of selected decision problems, understanding the significance of complexity classes and intractability.

Laboratory instructions or disclaimers if any:

Lab sessions will consist of solving numerical problems based on the algorithms discussed in theory lectures, discussions around finding complexities and analysis of algorithms.

(PCC) SQL and Python Programming

Teaching Scheme

Lectures: 3 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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

Course content

Database Design **[6 Hrs]**

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: **[6 Hrs]**

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 DatabaseFunctions: Numeric, Set Operations.

Advanced SQL: **[6 Hrs]**

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, Analyzing Data with Python, Working with Real world Dataset.

Basic Python Programming: **[6 Hrs]**

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: **[8 Hrs]**

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: **[6 Hrs]**

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 **[6 Hrs]**

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 **[6 Hrs]**

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.

Self-Study: **[12 Hrs]**

SQL Window Functions and Query Optimization, Stored Procedures and Triggers (Introduction), Python File Handling (CSV, JSON), Basic API Handling using Python, Mini Project using SQL and Python on a real-world dataset (Data analysis and visualization)

Textbooks

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

Reference Books

1. Rob Coronel, "Database systems: Design implementation and management", 4thEdition, Thomson Learning Press.
2. Richard L. Halterman, Learning to Program with Python.
3. Miller Curtis, Hands-On Data Analysis with NumPy and pandas, Packt Publishing Limited, ISBN: 9781789530797.
4. Jake Vanderplas, Python Data Science Handbook, O'Reilly Media, ISBN: 978-1-491-91205-8

Web references

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

Laboratory assignments

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. Create a schema using SQL DDL Commands and perform basic SQL DML commands on that schema.
2. Getting started with Jupyter Notebook, Implementing basic programming concepts.
3. Implementing Data structures and their functionalities.
4. Applying list comprehension and using lambda operators with map, filter and reduce functions to solve a scenario.
5. Analyzing Image Data using Numpy.
6. Perform Data Manipulation on a dataset using Panda.
7. Perform EDA (Exploratory Data Analysis) on a dataset using Panda.
8. Perform Data visualization using static and interactive plots on a given dataset.
9. Connect to a RDBMS and execute SQL queries on a dataset.

(PCC) Machine Learning

Teaching Scheme

Lectures: 3 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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.

Supervised Learning: **[6 Hrs]**

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

Classification Algorithms **[6 Hrs]**

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

Decision Tree **[6 Hrs]**

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

Dimensionality reduction and Feature Selection **[6 Hrs]**

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

Performance Measures **[6 Hrs]**

Performance metrics for classification, Thresholding, ROC and AUC curve

Self-study:

Dimensionality reduction: LDA; Ensemble methods: Random forest, boosting; Clustering: k-means, k-medoids, DBSCAN, Gaussian Mixture Models, Silhouette Score; Software libraries: NumPy, Scikit-Learn

Textbooks

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

Laboratory 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.

Laboratory instructions or disclaimers if any

(PCC) Data Engineering

Teaching Scheme

Lectures: 3 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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

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

Data Governance, security and privacy **[6 Hrs]**

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

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

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.

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

Self Study

Modern Data Stack Overview - ELT vs ETL; Data Mesh and Data Fabric – Concepts and Use Cases ; Introduction to Workflow Orchestration Tools - Airflow, Prefect ; Schema Evolution and Data Versioning ; Data Observability and Reliability Engineering ; Real-World Data Engineering Case Studies - Streaming vs Batch pipelines ; Ethical Data Usage and Responsible AI ; Industry Trends in Data Engineering.

Textbooks

1. Data Pipelines Pocket Reference: Moving and Processing Data for Analytics by James Densmore, OREILLY

Web references

1. Professional Data Engineer Certification exam guide, Google Cloud (https://cloud.google.com/certification/guides/data-engineer

Laboratory assignments

1. Pipeline for Data Engineering taking data from social media.
2. Data Preprocessing for the data downloaded from the Assignment No. 1
3. a. Installation of PySpark (latest version), data upload and perform mathematical operations.
   b. To upload Test file and get word count and Character count.
4. a. Installation of Hadoop (latest Version), creating account and Creating Directory.
   b. Uploading the text file and display the contents of uploaded file, Copy the file back from HDFS to Local and Remove the files for HDFS.
5. Data Visualization for the Data downloaded for Assignment No.1 using Seaborn and Matlpotlib library.
   1. histogram 2. bar chart
   3. scatter plot 4. Pie Chart
   5. Heat map 6. Pair Plot
   7. Violin
6. Data processing system: Implementation using Online (free) cloud platform - Looker Studio.

Laboratory instructions or disclaimers if any

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

Teaching Scheme

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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.

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.

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.

Linear Algebra **[6 Hrs]**

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

Textbooks

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

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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.

Organizing & Simplifying Data **[6 Hrs]**

Understand how to: Filter data, Sort data, Build groups, Build hierarchies, Build sets.

Field & Chart Types **[6 Hrs]**

Difference between measures and dimensions, difference between discrete and continuous fields,Tableau-generated fields, Understand how and when tobuild: 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.

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.

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.

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.

Dashboards **[6 Hrs]**

Build dashboards and stories, create dashboard actions, Design dashboards for viewing on devices, utilize visual best practices for viewing on devices, describe publishing & sharing options.

Textbooks

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

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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

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

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

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

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

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.

Textbooks

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

Reference Books

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

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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 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

Course content

[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 problems, data collection, analysis, interpretation, necessary instrumentations

[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

[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

[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

[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.

[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 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**
Lectures: 1 Hrs/ Week	Theory: MSE: 50 Marks, TA: 50 Marks
Laboratory: 2 Hrs/ Week	Laboratory: CIE: 100 Marks
Self-Study: 1Hrs/Week	

Course outcomes

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

Course content

Fundamentals of Communication **[4 Hrs]**

7 Cs of communication, common errors in English, enriching vocabulary, styles and registers

Aural-Oral Communication **[4 Hrs]**

The art of listening, stress and intonation, group discussion, oral presentation skills

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 references

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.html

[OE] Data Structures

Teaching Scheme

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

1. Understand and use the basic concepts for algorithm complexity
2. Use advanced data structures for efficient searching
3. Employ various data structures for implementing priority queues.
4. Analyse the time and space complexity string data structures.
5. Use different data structures to solve graph problems.
6. Design solutions for real-life problems using appropriate data structures

Course content

Algorithm Concepts **[8 Hrs]**

Abstract data types, Data structures, Algorithms, Big Oh, Small Oh, Omega and Theta notations, solving recurrence equations, Master theorems, Generating function techniques, Constructive induction.

Advanced Search Structures for Dicnary ADT **[8 Hrs]**

trees, Amortized analysis, 2-3 trees, 2-3-4 trees, Red-black trees, Randomized structures, Skip lists, Treaps, Universal hash functions.

Advanced Structures for Priority Queues **[6 Hrs]**

Binary Heap, Min Heap, Max Heap, Binomial heaps, Leftist heaps, Skewed heaps, Fibonacci heaps and its amortized analysis, Applications to minimum spanning tree algorithms.

Data Structures for Strings **[8 Hrs]**

Introduction to string data structures, Huffman coding tree, Tries, Compressed Tries, Suffix Trees, Suffix Arrays; Applications-Search Engines, Bioinformatics, Pattern Matching: KMP algorithm; Internet Packet Forwarding.

[6 Hrs]

Graph Algorithms

BFS, DFS, Disjoint set union problem, Network flow; Maximum-Flow / Minimum-Cut; Ford–Fulkerson algorithm, Hamiltonian Path and circuit problem, Introduction to Hypergraphs.

Self-study

Geometric data structures, Cut vertices, Plane sweep paradigm, Hamiltonian Path and circuit problem.

Textbooks

1. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, Introduction to Algorithms, 3rd Edition, PHI Learning Pvt. Ltd.; ISBN-13: 978-0262033848 ISBN-10: 0262033844

2. Robert Sedgewick and Kevin Wayne, Algorithms, Pearson Education, 4th Edition, ISBN-13: 978-0321573513

Reference Books

1. S. Dasgupta, C.H. Papadimitriou, and U. V. Vazirani, Algorithms, McGraw-Hill, 2006; ISBN- 13: 978-0073523408 ISBN-10: 007352340, 2
2. J. Kleinberg and E. Tardos, Algorithm Design, Addison-Wesley, 2006; ISBN-13: 978- 0321295354 ISBN-10: 0321295358x`

Big Data Analytics with Apache Spark

Teaching Scheme

Lectures: 3 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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

Course content

Introduction to Spark **[6 Hrs]**

Introduction to Apache Hadoop, Overview of Hadoop Ecosystem, Spark – Introduction, Sparsk – Ecosystem Components. Spark basics: Features and Use Cases, SparkContext, Stage, Executor.

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.

Spark SQL AND dataframe **[8 Hrs]**

Spark SQL – Introduction, Spark SQL – Features, Spark SQL – DataFrame, Spark SQL – DataSet, Spark SQL – Optimization, HIVE Fundamentals

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.

KAFKA **[6 Hrs]**

Kafka – Introduction, Kafka Architecture, Kafka Workflow, Kafka – Cluster configuration, Kafka monitoring tools.

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.

Self-Study:

Spark Internals - Job Execution Flow (High-level) ; Catalyst Optimizer and Tungsten Engine; Advanced Spark SQL Use Cases ; Integration of Spark with Cloud Platforms - AWS EMR, Azure Databricks ; Spark Performance Anti-

Patterns ; Real-Time Analytics - Use Cases with Spark Streaming ; Introduction to Delta Lake and Lakehouse Architecture ; Case Studies - Spark in Industry (Finance, IoT, E-commerce)

Textbooks

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

Laboratory 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

Lectures: 3 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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

Course content

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.

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

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)

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)

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

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

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

Textbooks

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

Laboratory 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.

MLOps and Systems

Teaching Scheme

Lectures: 3 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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)

Course content

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.

ML API Development and Deployment on GCP and AWS [6 Hrs]

Flask, FastApi, Tensorflow serving, Tensorflow lite for optimization latency

Cloud-based APIs [6 Hrs]

Intro to Kubeflow, Collaborative platforms: MLflow, AutoML in GCP, Azure, AWS, Hyper parameter optimization

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

Application Production Testing on Google Optimize [6 Hrs]

Canary pattern, A/B Pattern, Shadow Pattern

Production Software Build and Analytics [8 Hrs]

Prometheus, Grafana Cloud, Cloudwatch, Special Topics : Building models from scratch, Hyper- parameterization, optimization, quantization

Self-Study:

A comparative study of two orchestration stacks (e.g., Kubeflow vs. Prefect) on the same pipeline. A production case study (from an open-source repo or paper) dissecting deployment, monitoring, and rollback. An LLMOps mini project: prompt/versioning, evaluation harness, and a safe rollout plan.

Textbooks

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

Laboratory 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

Laboratory instructions or disclaimers if any

DS (DE1) Time Series Data Analysis

Teaching Scheme

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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

Basic Forecasting Models **[8 Hrs]**

Naïve Approach, Moving Average Method, Simple exponential Smoothing method, Holt's Linear Trend Model, Holt's Winter Model.

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.

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.

Textbooks

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

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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.

Self-study **[6 Hrs]**

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

Textbooks

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

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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

Course content

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

Data Structures and Data Handling in R [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()

Control Structures, Functions and Object-Oriented Concepts in R **[6 Hrs]**

Conditions and loops, control functions, apply family, default values for arguments, lazy evaluation in functions, objects and classes, debugging

Data Analysis, Visualization and Machine Learning using 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

Self Study: Practice on built-in datasets in R, additional exercises on vectors, matrices and data frames, writing user-defined functions, hands-on practice on data import/export, data cleaning, exploratory data analysis, and basic statistical and machine learning implementations in R.

Textbooks

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

Lectures: 3 Hrs/ Week

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

Introduction to GANs **[8 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

Deep Convolutional GAN **[8 Hrs]**

Convolutional neural networks, Convolutional filters, Parameter sharing, Batch normalization, Training challenges (Mode collapse, Slow convergence, Overgeneralization), Non-Saturating GAN

Various types of GANs **[14 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

Real-World Applications for GANs **[10 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

Self Study

Reviewing the latest research papers, Case studies on Deepfake detection and mitigation, medical images for data augmentation, Image-to-image translation for style transfer and creative applications etc.

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

Laboratory: 2 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Laboratory: CIE: 50 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

Introduction to Reinforcement Learning **[4 Hrs]**

Examples, Elements of Reinforcement Learning, Limitations and Scope, An Extended Example: Tic-Tac-Toe, Early History of Reinforcement Learning

Multi-armed Bandits **[8 Hrs]**

k-a rmed 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

Finite Markov Decision Processes **[8 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

Dynamic Programming (DP) **[8 Hrs]**

Policy Evaluation (Prediction), Policy Improvement, Policy Iteration, Value Iteration, Asynchronous DP, Generalized Policy Iteration Efficiency of DP

Monte Carlo (MC) Methods **[8 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

Temporal-Difference (TD) Learning **[8 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

Simple Mini Self-Study Projects

1. Implement ε-greedy bandit (5 lines logic).
2. Implement simple Gridworld with random rewards.
3. Compare SARSA and Q-learning on same problem.
4. Plot reward vs episodes graph.

Textbooks

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

Lectures: 3 Hrs/ Week

Self-Study: 1Hrs/Week

Evaluation Scheme

Theory: MSE: 30 Marks, TA: 20 Marks, ESE: 50 Marks

Course outcomes

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.

Course content

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.

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.

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.

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.

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.

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.

Self-study:

Transformer, Encoder, decoder, attention mechanism, multi-head attention, positional encoding, optimization measures, Application of word embedding to shallow Parsing- Morphological Processing, Part of speech Tagging and chunkling, sequence to sequence (seq2seq) transformation using deep learning : LSTMs and Variants

Textbooks

1. Daniel Jurafsky and James H. Martin, "Speech and Language Processing", Second Edition,
2. Prentice Hall, 2008, ISBN: 978-0131873216.
3. Allen James, "Natural Language Understanding", Second Edition, Benjamin/Cumming, 1994, ISBN: 978-0805303346.
4. 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).

Detailed Syllabus: Semester III

[SLC] Massive Open Online Course – I

Teaching Scheme
Self-Study: 3hours / Week

Examination Scheme
CIE: 50 Marks, ESE: 50 marks

Course Outcome
Students will be able to:

1. Acquire new skills or knowledge to enhance their personal and professional development
2. Receive a flexible learning environment, allowing one to study at own pace and convenience
3. Opporty for lifelong learning
4. Foster collaboration and networking among participants

The students in consultation with the faculty advisor, opt for a single course of 12 weeks offered by the NPTEL in the current semester. The students need to register for the examination conducted by the NPTEL. For the students who secured a passing score in the NPTEL examination, the marks obtained for assignments (in 25 marks) will be upscaled to out of 50 marks of CIE and the marks obtained from the certificate examination (in 75 marks) will be downscaled 50 marks of ESE assessments.

[SLC] Massive Open Online Course – II

Teaching Scheme
Self-Study: 3 hours / Week

Examination Scheme
CIE: 50 Marks, ESE: 50 marks

Course Outcome
Students will be able to:

1. Acquire new skills or knowledge to enhance their personal and professional development
2. Receive a flexible learning environment, allowing one to study at own pace and convenience
3. Opporty for lifelong learning
4. Foster collaboration and networking among participants

The students in consultation with the faculty advisor opt for a single course of 12 weeks offered by the NPTEL in the current semester. The students need to register for the examination conducted by the NPTEL. For the students who secured a passing score in the NPTEL examination, the marks obtained for assignments (in 25 marks) will be upscaled to out of 50 marks of CIE and the marks obtained from the certificate examination (in 75 marks) will be downscaled 50 marks of ESE assessments.

[Project] Dissertation Phase – I

Teaching Scheme	Examination Scheme
Laboratory: 22 hours/week	Theory: CIE: 70 Marks
Self-Study: 12 hours / Week	ESE: 30 marks

Course Outcomes

Students will be able to:

1. Demonstrate how to search the existing literature to gather information about a specific problem or domain.
2. Identify the state-of-the-art technologies and research in the chosen domain and highlight open problems that are relevant to societal or industrial needs.
3. Evaluate various solution techniques to determine the most feasible solution within the given constraints for the chosen dissertation problem.
4. Apply software engineering principles related to requirements gathering and design to produce relevant documentation.
5. Write a dissertation report that details the research problem, objectives, literature review, and solution architecture.
6. Deliver effective oral presentations to communicate the findings and outcomes of the research work.

Guidelines

The dissertation is a year-long project, conducted and evaluated in two phases. It can be carried out either in-house or within an industry as assigned by the department. The project topic and internal advisor (a faculty member from the department) are determined at the beginning of Phase I.

Students are expected to complete the following activities in Phase-I:

1. Literature survey
2. Problem Definition
3. Motivation for study and Objectives
4. Preliminary design /feasibility / modular approaches

Deliverables

1. A report having the following details: Abstract, Problem statement, Requirements specification, Literature survey, Proposed solution, High-level design description, Plan for implementation and testing in Phase-II
2. A presentation that covers the major points covered in the report.
3. A proof of concept (preferably, but not mandatory)

Evaluation

Two independent assessments (Mid-Semester and End-Semester evaluations) will be made. In both the Examinations, the internal guide, along with a Senior Faculty member of the department, will evaluate the work. The marks obtained in these two assessments will be combined to get the final evaluation out of 100 marks. The course grading, like other courses, will be relative in nature.

The evaluation will take place based on criteria such as literature survey and well-defined project problem statement, proposed high level system design, concrete plan for implementation and result generation, presentation etc.

The panel (external examiner(s) and senior faculty) will provide a report about suggestions/changes to be incorporated during phase-II.

Detailed Syllabus: Semester IV

[VSEC] Dissertation Phase – II

Teaching Scheme	**Examination Scheme**
Laboratory: 22 hours/week	Theory: CIE: 70 Marks
Self-Study: 12 hours / Week	ESE: 30 marks

Course Outcomes

Students will be able to:

1. Achieve proficiency in the languages, tools, libraries, and technologies used in the dissertation work.
2. Apply project planning principles and techniques to ensure effective and efficient project execution.
3. Demonstrate an understanding of the entire lifecycle of a software product or solution.
4. Produce artifacts such as source code, test plans, and test results based on the dissertation work.
5. Write research paper(s) and a thesis in accordance with publication ethics.
6. Exhibit the presentation skills needed to effectively present the work at various platforms.

Guidelines

Student is expected to complete the following activities in Phase-II:

1. Implementation of the proposed approach in the first stage
2. Testing and verification of the implemented solution
3. Writing of a report and presentation
4. Publish the work done at a suitable Scopus indexed conference/in a journal

Deliverables

1. Source code (if the project is in-house)
2. Dissertation report that gives overview of the problem statement, literature survey, design, implementation details, testing strategy and results of testing
3. All the artifacts created throughout the duration of dissertation such as requirements specification, design, project plan, test cases etc
4. Presentation based on the dissertation report
5. Research Paper(s) based on the dissertation work

Evaluation

Evaluation will be done in two steps: Mid-Semester evaluation and End-Semester evaluation. In the Mid Semester Examination, the internal guide, along with a Senior Faculty of the department, will evaluate the work. In the End Semester Examination evaluation, the internal guide, along with an external expert (usually from an Industry) will evaluate the work. The marks obtained in these two assessments will be combined to get the final evaluation out of 100 marks. The course grading, like other courses, will be relative in nature.

The assessment is done on the criteria such as concrete system design, implementation status and concrete plan for completion of remaining tasks, presentation etc.

The purpose of Mid-Semester evaluation is also to check preparedness of students for the End- Semester evaluation. Examiners may give suggestions for changes/corrections to be incorporated before the final evaluation. If the work done till then may not lead to successful completion of the dissertation in the remaining time, the student may be asked to take an extension in time to complete the course.

The assessment End-Semester evaluation will be done based on the criteria such as quality of implementation, result analysis, project outcomes (publications, patent, copyright, contribution to open-source commy, participation in project competition etc.), quality of report, presentation etc.

The total assessment of phase-II work is for 100 marks and the grading, like other courses, will be relative.