Savitribai Phule Pune University, Pune

Maharashtra, India

Faculty of Science and Technology

Curriculum Structure and Syllabus

Master of Engineering (2025 Pattern) in

M. E. - Data Science (DS)

(With effect from Academic Year 2025-26)

www.unipune.ac.in

Contents

Nomenclature

CCE	Continuous Comprehensive Evaluation
ESE	End Semester Examination
KAP	A Knowledge and Attitude Profile
L	Lecture
OR	Oral
PCC	Programme Core Course
PEC	Programme Elective Course
PEO	Programme Educational Objectives
PO	Programme Outcomes
PSO	Program Specific Outcomes
TW	Term Work
WK	Knowledge and Attitude Profile

Preface by Board of Studies

Dear Teachers and Students,

On behalf of the Board of Studies in Computer Engineering, we are pleased to introduce the revised syllabus for the **Master of Engineering (Data Science)** program, effective from the academic year 2025–26. This comprehensive and forward-thinking curriculum has been thoughtfully crafted to equip students with a strong foundation in data science principles while ensuring hands-on exposure to real-world data challenges across diverse sectors.

As data becomes the cornerstone of decision-making in todays digital economy, our program aims to develop proficient data scientists who can derive meaningful insights, build scalable models, and drive data-driven innovation. The syllabus encompasses a wide spectrum of essential topics including Machine Learning, Big Data Technologies, Cloud Computing, Data Visualization, and Data Governance, along with a strong emphasis on ethics, privacy, and responsible data usage.

We would like to express our heartfelt appreciation to the dedicated faculty members, industry professionals, academic researchers, and students who have contributed to this syllabus revision. Their insights and collaborative efforts have ensured that the program stays relevant to evolving industry needs and global trends in data science and analytics. We are confident that this newly updated syllabus will serve as a robust academic framework that not only strengthens technical expertise but also nurtures critical thinking, problem-solving, and ethical data stewardship.

Dr. Nilesh Uke
Chairman
Board of Studies - Computer Engineering.

Members of Board of Studies

Dr. Pramod Patil	Dean (S&T) , Dr. D.Y. Patil Institute of Technology, Pune
Dr. Amol Potgantwar	Sandip Institute of Technology and Research Center, Nasik
Dr. Sachin Babar	Sinhhgad Institute of Technology Lonavala
Dr. Dhananjay Kshirsagar	Sanjivani College of Engineering, Kopargaon
Dr. Balwant Sonkamble	Pune Institute of Computer Technology, Pune
Dr. Suhasini Itkar	PES's Modern College of Engineering, Pune
Dr. Dipti Patil	Cummins College of Engineering for Women, Pune
Dr. Sachin Sakhare	Vishwakarma Institute of Information Technology, Pune
Dr. Dipak Patil	R.H. Sapat college of Engineering, Nasik
Dr. Vandana Dhingra	Department of Technology, Savitribai Phule Pune University
Dr. Deepali Ujalambkar	AISSMS College of Engineering, Pune
Dr. Vaishali Vikhe	Pravara Rural Engineering College, Loni
Dr. Pradip Jawandhiya	PR Pote College of Engineering and Management, Amravati
Dr. Sandeep Deshmukh	TeemGenie, Pune

Programme Educational Objectives (PEO)

Program education objectives are broad statements that describe the career and professional accomplishments that the program is preparing graduates to achieve.

PEO	PEO Focus	PEO Statements
PEO1	Technical Expertise	To equip postgraduates with advanced knowledge and practical skills in data science, including data engineering, statistical modeling, machine learning, and big data analytics, enabling them to design and implement data-driven solutions across diverse sectors such as healthcare, finance, marketing, and smart cities.
PEO2	Ethical & Societal Impact	To prepare postgraduates to consider ethical, legal, and societal implications of data collection, analysis, and usage, promoting responsible data practices that uphold privacy, fairness, and equity in decision-making.
PEO3	Research & Innovation	To nurture a research-oriented mindset among postgraduates, encouraging them to pursue innovation in data science methodologies, contribute to open research, and collaborate across disciplines to address real-world problems.
PEO4	Leadership & Collaboration	To develop postgraduates into effective leaders and communicators who can manage data science projects, articulate insights to both technical and non-technical audiences, and work collaboratively in multidisciplinary environments to drive data-centric organizational and societal progress.

Knowledge and Attitude Profile (WK)

A Knowledge and Attitude Profile (KAP), often represented as WK (Knowledge and Attitude Profile) in some contexts, is a framework or assessment tool used to evaluate an individual's knowledge and attitudes related to a specific area, topic, or domain.

WK1	A systematic, theory-based understanding of the natural sciences applicable to the discipline and awareness of relevant social sciences.
WK2	Conceptually-based mathematics, numerical analysis, data analysis, statistics and formal aspects of computer and information science to support detailed analysis and modelling applicable to the discipline.
WK3	A systematic, theory-based formulation of engineering fundamentals required in the engineering discipline.
WK4	Engineering specialist knowledge that provides theoretical frameworks and bodies of knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline.
WK5	Knowledge, including efficient resource use, environmental impacts, whole-life cost, re-use of resources, net zero carbon, and similar concepts, that supports engineering design and operations in a practice area.
WK6	Knowledge of engineering practice (technology) in the practice areas in the engineering discipline.
WK7	Knowledge of the role of engineering in society and identified issues in engineering practice in the discipline, such as the professional responsibility of an engineer to public safety and sustainable development.
WK8	Engagement with selected knowledge in the current research literature of the discipline, awareness of the power of critical thinking and creative approaches to evaluate emerging issues.
WK9	Ethics, inclusive behavior and conduct. Knowledge of professional ethics, responsibilities, and norms of engineering practice. Awareness of the need for diversity by reason of ethnicity, gender, age, physical ability etc. with mutual understanding and respect, and of inclusive attitudes.

Master of Engineering - Data Science (2025 Course)

Programme Outcomes (PO)

Program Outcomes are statements that describe what students are expected to know and be able to do upon graduating from the program. These relate to the skills, knowledge, attitude and behaviour that students acquire through the program. NBA has defined the following three POs for a graduate of PG Engineering Program:-

PO1	Ability to independently carry out research/investigation and development work to solve practical problems.
PO2	Ability to write and present a substantial technical report/document.
PO3	Ability to demonstrate a degree of mastery over the area of specialization (Computer Engineering) at a level higher than the bachelor's program.

Program Specific Outcomes (PSOs)

Program Specific Outcomes (PSOs) are statements that describe the knowledge, skills, and attitudes that graduates of a academic program (Master of Engineering in Data Science) should be able to demonstrate at the time of their graduation.

PSO1	**Data Analytics and Modeling:** Apply advanced statistical, computational, and machine learning techniques to analyze complex datasets and develop predictive/prescriptive models for decision-making.
PSO2	**Data Engineering and System Development:** Design, develop, and deploy scalable, secure, and efficient data-driven systems using big data technologies, cloud platforms, and modern data engineering practices.
PSO3	**Research, Innovation, and Application:** Conduct research, evaluate emerging technologies, and innovate data science solutions for interdisciplinary domains such as business, healthcare, finance, and social sciences.

Curriculum Structure

First Year – Master of Engineering in Data Science (2025 Pattern)

Semester I

Course Code	Course Type	Course Name	Teaching Scheme (Hours/ Week)		Examination Scheme and Marks						Credits		
			L	P	CCE	ESE	TW	PR	OR	Total	L	P	Total
PCC-501-DS	Programme Core Course	Mathematical Foundations	4	-	50	50	-	-	-	100	4	-	4
PCC-502-DS	Programme Core Course	Big Data Technologies	4	-	50	50	-	-	-	100	4	-	4
PCC-503-DS	Programme Core Course	Machine Learning	4	-	50	50	-	-	-	100	4	-	4
PCC-504-DS	Programme Core Corse	Advanced Data Analysis and Visualization	4	-	50	50	-	-	-	100	4	-	4
PCC-505-DS	Programme Core Course	Computational Laboratory - I	-	4	-	-	25	-	25	50	-	2	2
PEC-521-DS	Programme Elective Course	Elective-1	3	-	50	50	-	-	-	100	3	-	3
PEC-522-DS	Programme Elective Course	Skill Based Laboratory - 1	-	2	-	-	25	-	25	50	-	1	1
Total			**19**	**6**	**250**	**250**	**50**	**-**	**50**	**600**	**19**	**3**	**22**

List of Elective I Courses	
PEC-521A-DS	Optimization Techniques
PEC-521B-DS	Advanced Information Retrieval
PEC-521C-DS	Quantum Computing
PEC-521D-DS	Computer Vision

First Year – Master of Engineering in Data Science (2025 Pattern)

Semester II

Course Code	Course Type	Course Name	Teaching Scheme (Hours/ Week)		Examination Scheme and Marks						Credits		
			L	P	CCE	ESE	TW	PR	OR	Total	L	P	Total
PCC-551-DS	Programme Core Course	Data Engineering	4	-	50	50	-	-	-	100	4	-	4
PCC-552-DS	Programme Core Course	Cloud Computing for Data Analytics	4	-	50	50	-	-	-	100	4	-	4
PCC-553-DS	Programme Core Course	Deep Learning	4	-	50	50	-	-	-	100	4	-	4
PCC-554-DS	Programme Core Course	Computational Laboratory - II	-	4	-	-	25	-	25	50	-	2	2
PEC-561-DS	Programme Elective Course	Elective-II	3	-	50	50	-	-	-	100	3	-	3
PEC-562-DS	Programme Elective Course	Elective-III	3	-	50	50	-	-	-	100	3	-	3
SEM-571-DS	Seminar	Technical Seminar I	-	4	-	-	25	-	25	50	-	2	2
Total			**18**	**8**	**250**	**250**	**50**	**-**	**50**	**600**	**18**	**4**	**22**

List of Elective II Courses		List of Elective III Courses	
PEC-561A-DS	Generative AI	PEC-562A-DS	Financial Analytics
PEC-561B-DS	Blockchain Technology	PEC-562B-DS	Healthcare Analytics
PEC-561C-DS	Web Mining and Analytics	PEC-562C-DS	Social Network Analytics
PEC-561D-DS	Data Privacy and Ethics	PEC-562D-DS	IoT Analytics

Second Year – Master of Engineering in Data Science (2025 Pattern)

Semester III

Course Code	Course Type	Course Name	Teaching Scheme (Hours/ Week)		Examination Scheme and Marks						Credits		
			L	P	CCE	ESE	TW	PR	OR	Total	L	P	Total
RM-601-DS	Research Methodology	Research Methodology	4	-	50	50	-	-	-	100	4	-	4
OJT-602-DS	OJT /Internship	On Job Training/ Internship	-	10	-	-	100	-	-	100	-	5	5
SEM-603-DS	Seminar	Technical Seminar II	-	8			25	-	25	50	-	4	4
RPR-604-DS	Research Project	Research Project - Stage I	-	18	-	-	25	-	25	50	-	9	9
Total			**4**	**36**	**50**	**50**	**150**	**-**	**50**	**300**	**4**	**18**	**22**

Semester IV

Course Code	Course Type	Course Name	Teaching Scheme (Hours/ Week)		Examination Scheme and Marks						Credits		
			L	P	CCE	ESE	TW	PR	OR	Total	L	P	Total
SEM-651-DS	Seminar	Technical Seminar III	-	8	-	-	50	-	50	100	-	4	4
RPR-652-DS	Research Project	Research Project - Stage II	-	36	-	-	150	-	50	200	-	18	18
Total			**-**	**44**	**-**	**-**	**200**	**-**	**100**	**300**	**-**	**22**	**22**

Savitribai Phule Pune University, Pune

Maharashtra, India

M. E. - Data Science (DS) - 2025 Pattern

Semester I

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-501-DS - Mathematical Foundations		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks Hours **End-Semester**: 50 Marks

Course Objectives: The course aims to:

1. To build a strong mathematical foundation essential for understanding and designing AI and Machine learning algorithms.
2. To apply key concepts from linear algebra, probability, statistics, calculus, and graph theory in AI contexts.
3. To explore optimization techniques critical to model training and performance improvement.
4. To develop the ability to analyze, model, and interpret data-driven AI systems using mathematical reasoning.

Course Outcomes: Upon successful completion of this course, students will be able to:

- CO1: Develop mathematical rigor for cutting-edge technologies.
- CO2: Utilize mathematics as a tool to interpret the outcomes of the black box techniques.
- CO3: Develop optimization techniques for learning algorithms.
- CO4: Apply statistical techniques to model the data.

Course Contents

Unit I - Linear Algebra - (12 Hours)

Introduction to scalars, vectors, matrices, and tensors, Matrix operations: addition, multiplication, transposition, Properties of matrices: rank, inverse, determinant, Systems of linear equations and their solutions, Eigenvalues and Eigenvectors: significance in PCA and dimensionality reduction, Singular Value Decomposition (SVD) and its applications in recommender systems

Unit II Calculus and Optimization (12 Hours)

Gradient vector and its interpretation, Directional derivatives, Tangent planes and linear approximation, Critical points and saddle points, Second derivative test (Hessian), Constrained optimization using Lagrange multipliers

Unit III Probability (12 Hours)

Probability mass, density, and cumulative distribution functions, Parametric families of distributions, Expected value, variance, conditional expectation, Applications of the univariate and multivariate, Central Limit Theorem, Probabilistic inequalities, Bayes Theorem, Markov chains

Unit IV Statistics (12 Hours)

Descriptive Statistics: Mean, Median, Mode, Range, Variance, Standard Deviation, Skewness, Kurtosis, Data Visualization: Histograms, Bar plots, Pie charts, Introduction to Statistical Inference Problems; Point Estimation; Interval Estimation; Testing of Hypotheses; Two Sample Problems Involving Normal Populations, Tests for Proportions, Chi Square Goodness of Fit Test, Contingency Tables.

Unit V Discrete Mathematics and Graph Theory (12 Hours)

Mathematics: Sets, Relations, functions, Probability Theory: P and C, GraphTheory: Isomorphism, Planar graphs, graph coloring, Hamilton circuits and Euler cycles. Permutations and Combinations with and without repetition. Specialized techniques to solve combinatorial enumeration problems.

Learning Resources

Text Books:

1. M. P. Deisenroth, A. A. Faisal, and C. S. Ong, Mathematics for Machine Learning. Cambridge, U.K.: Cambridge University Press, 2020.
2. E. Kreyszig, K. Stroud, and G. Stephenson, Advanced engineering mathematics, Integration, vol. 9, no. 4, p. 1014, 2008.
3. W. L. Briggs, L. Cochran, B. Gillett, and E. Schulz, Multivariable Calculus, 3rd ed. Boston, MA: Pearson Education, 2018.
4. S. Boyd and L. Vandenberghe, Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares. Cambridge, U.K.: Cambridge University Press, 2018.

Reference Books:

1. C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY: Springer, 2006.
2. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.
3. G. James and P. Dyke, Advanced Modern Engineering Mathematics, 5th ed. Harlow, U.K.: Pearson Education Limited, 2018.
4. S. Ross, A First Course in Probability, 9th ed. Boston, MA: Pearson, 2012.

NPTEL/MOOC/Youtube Links:

1. Foundational Mathematics for AI, Coursera. [Online]. Available: https://www.coursera.org/learn/fou mathematics-for-ai.
2. Study Materials | Linear Algebra | Mathematics, MIT OpenCourseWare. [Online]. Available: https://ocw.mit.edu/courses/18-06-linear-algebra-spring-2010/pages/study-materials/.

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-502-DS - Big Data Technologies		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses :

Foundation in Distributed systems, Programming, Databases, Statistics and basic linear algebra, Cloud Basics

Course Objectives: The course aims to:

1. To understand the challenges and architecture of large-scale data processing.
2. To explore the big data ecosystem and its components.
3. To provide hands-on experience with distributed data frameworks and analytics tools.

Course Outcomes: Upon successful completion of this course, students will be able to:

- CO1: Identify the characteristics, architecture, and challenges of Big Data systems.
- CO2: Implement distributed storage and processing using Hadoop, HDFS, and MapReduce.
- CO3: Analyze data using Spark, Hive, and other big data analytics engines.
- CO4: Differentiate between batch and stream processing frameworks like Spark Streaming and Flink.
- CO5: Design scalable big data solutions using cloud-based big data services (EMR, Dataproc).

Course Contents

Unit I -Introduction to Big Data - (12 Hours)

Definition and characteristics (5Vs); Big data vs traditional data; Applications in IoT, healthcare, finance, etc.; Challenges; Distributed computing; Cloud-based storage (S3, BigQuery).

Unit II - Big Data Ecosystem (12 Hours)

Hadoop: HDFS, MapReduce, YARN; Spark fundamentals; Pig, Hive, and HBase; ZooKeeper and Sqoop; Architecture and configuration.

Unit III - Distributed and Stream Processing Frameworks - (12 Hours)

Apache Spark vs Flink; Batch vs real-time; Kafka, Storm, Kinesis; Use cases in event processing and real-time analytics.

Unit IV - Big Data Analytics and Query Engines (12 Hours)

DataFrames and Datasets in Spark; Structured streaming; HiveQL; Presto, Druid, Clickhouse; Integration with BI tools.

Unit V - Big Data in the Cloud & Governance (12 Hours)

Big data as a service: EMR, Dataproc, HDInsight; Data governance: Quality, lineage, cataloging; Security and privacy in large-scale systems.

Learning Resources

Text Books:

1. N. Marz, Big Data: Principles and Best Practices of Scalable Real-Time Data Systems. Manning Publications, 2015.
2. T. White, Hadoop: The Definitive Guide, 4th ed. O'Reilly Media, 2015.

Reference Books:

1. J. S. Damji, B. Das, D. Wenig, and T. Dunning, Learning Spark: Lightning-Fast Big Data Analysis, 2nd ed. O'Reilly Media, 2020.
2. V. Lakshmanan, Data Science on the Google Cloud Platform. O'Reilly Media, 2018

SWAYAM / MOOC / YouTube Links

1. R. Misra, Big Data Computing, NPTEL, IIT Patna. [Online]. Available: https://nptel.ac.in/courses/106
2. Big Data Specialization, Coursera. [Online]. Available: https://www.coursera.org/specializations/big-data

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-503-DS - Machine Learning		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses : Students should have prior knowledge of Fundamentals of Programming Languages, Design and Analysis of Algorithms

Course Objectives: The course aims to:

1. To understand the foundational concepts, types, and goals of Machine learning, including model evaluation and generalization strategies.
2. To explore various Supervised and Unsupervised learning algorithms and their practical applications.
3. To apply feature selection, model selection, and evaluation techniques to assess algorithm performance.
4. To analyze Ensemble learning methods and optimization techniques for improving predictive models.
5. To examine evolutionary algorithms and the fundamentals of Neural Networks and Computational learning theory

Course Outcomes: After successful completion of the course, Students will be able to:

- CO1: Develop critical thinking, and problem-solving ability using Machine learning tools.
- CO2: Apply Machine learning to assist in data-driven decision making.
- CO3: Develop mathematical background for Machine learning frameworks.
- CO4: Optimize the model with Regularization and Hyperparameter tuning

Course Contents

Unit I - Introduction to Machine Learning Concepts - (12 Hours)

Knowledge based systems, Rule based systems, Introduction to Machine Learning: Basic definitions, types of learning, Hypothesis space and inductive bias, Data pre-processing, handling data imbalance, Feature selection and extraction techniques, Dimensionality Reduction: PCA, t-SNE, LDA

Unit II - Supervised Learning Algorithms (12 Hours)

Linear Regression and Logistic Regression, k-Nearest Neighbors (k-NN), Decision Trees and Random Forests, Support Vector Machines (SVM), Model selection, Evaluation metrics (Accuracy, Precision, Recall. ROC, AUC), Overfitting, underfitting and generalization, L1 and L2 errors, Regularization: L1 and L2 regularization, Model evaluation: training/test/validation, cross-validation.

Unit III - Unsupervised Learning Algorithms - (12 Hours)

Clustering: k-Means, Hierarchical Clustering, DBSCAN, Evaluation metrics for clustering (Silhouette Score, Davies–Bouldin index, Adjusted Rand Index), Association Rule Mining: Apriori, FP-Growth.

Unit IV - Ensemble Methods and Model Optimization (12 Hours)

Bagging, Boosting, and Stacking, AdaBoost, Gradient Boosting, XGBoost, LGBM, CatBoost, Hyperparameter tuning: Grid Search, Random Search, Bayesian Optimization.

Unit V - Foundations and Frontiers in Machine Learning - (12 Hours)

Interpretability and Explainability: Need for explainability in ML, Global vs local interpretability, Tools and techniques: LIME, SHAP, Feature Importance, Trade-offs: Accuracy vs Interpretability, AutoML and Meta-Learning, Brief overview of tools: Auto-sklearn, H2O, Google AutoML, Meta-learning concepts (learning to learn).

Learning Resources

Text Books:

1. T. M. Mitchell, Machine Learning, 1st ed. New York, NY, USA: McGraw-Hill, 1997.
2. E. Alpaydin, Introduction to Machine Learning, 4th ed. Cambridge, MA, USA: MIT Press, 2020.
3. M. P. Deisenroth, A. A. Faisal, and C. S. Ong, Mathematics for Machine Learning. Cambridge, U.K.: Cambridge University Press, 2020.

Reference Books:

1. C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York, NY, USA: Springer, 2006.
2. K. P. Murphy, Machine Learning: A Probabilistic Perspective, 1st ed. Cambridge, MA, USA: MIT Press, 2012.
3. A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.
4. T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.
5. P. Kulkarni, Reinforcement and Systemic Machine Learning for Decision Making. Hoboken, NJ, USA: Wiley-IEEE Press, 2012.

SWAYAM / MOOC / YouTube Links

1. IBM, "IBM Machine Learning Professional Certificate," Coursera. [Online]. Available: https://www.cou certificates/ibm-machine-learning.
2. S. Sarkar, "Introduction to Machine Learning," NPTEL, IIT Kharagpur. [Online]. Available: https://onlinecourses.nptel.ac.in/noc21_cs85/preview.
3. A. Ng, "Machine Learning Specialization," Stanford Online. [Online]. Available: https://online.stanfor ymls-machine-learning-specialization.

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-504-DS - Advanced Data Analysis and Visualization		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks **End-Semester**: 50 Marks

Pre-requisite Courses : Basic programming, Data structure, Statistics

Course Objectives: The course aims to:

1. To master the fundamental principles and diverse techniques of data visualization for various data types and analytical tasks.
2. To gain hands-on proficiency in using industry-standard programming libraries and dedicated tools to implement static and interactive data visualizations and dashboards.
3. To develop effective data storytelling skills, transforming complex data into compelling narratives and actionable insights.
4. To critically evaluate data visualizations for effectiveness, clarity, potential biases, and ethical implications.
5. To design and implement robust, interactive data visualization solutions for real-world engineering problems.

Course Outcomes: Upon successful completion of this course, students will be able to:

- CO1: Design and implement effective static and interactive data visualizations for various data types
- CO2 : Applying principles of human perception and visual design using programming libraries and specialized tools.
- CO2: Develop comprehensive and interactive dashboards that facilitate exploratory data analysis and communicate key insights to diverse audiences.
- CO3: Construct compelling data stories by integrating visualizations with narrative elements, effectively conveying complex information and influencing decision-making.
- CO4: Critically evaluate the quality, effectiveness, and ethical implications of data visualizations, identifying potential misrepresentations or misleading designs.

Course Contents

Unit I - Foundations of Data Visualization and Design Principles - (12 Hours)

Introduction to Data Visualization, Definition, Role in Data Science and Engineering, Goals & Benefits, Human Perception & Cognition, Preattentive Attributes, Gestalt Principles, Visual Encoding, Marks & Channels, Effectiveness, Cognitive Load & Simplicity, Data Types & Analytical Tasks, Choosing the right visualization, Case Studies on Misleading Data Visualizations, Principles of Effective

Visualization Design, Clarity, Accuracy, Efficiency, Integrity, Ethical Visualization Practices, Avoiding Manipulative Designs.

Unit II - Static Visualization Techniques - (12 Hours)

Comparative Visualizations, Histograms, Box, Violin, Bar Charts, Scatter Plots, Line Charts, Bubble Charts, Stacked Area Charts, Geospatial Data Visualization, GIS techniques, Choropleth Maps, Symbol Maps, Graph & Network Visualizations, Tree Diagrams, Force-Directed Graphs, Animated Data Visualization, Introducing motion for enhanced storytelling, Case Study, Industry Applications of Static Visualizations

Unit III - Programming for Data Visualization - (12 Hours)

Python-Based Visualization, Matplotlib, Seaborn, Plotly, Bokeh, Data Preparation, Cleaning, Aggregation, Transformation, Real-Time Data Handling, Streaming visualization (e.g., Dash, WebSockets), Web-Based Visualization, SVG, Canvas, D3.js Introduction, Interactive Web-based Storytelling, Hands-on project.

Unit IV - Interactive Dashboards and Tools - (12 Hours)

Dashboard Design Principles, UX/UI Concerns, Navigation, Filtering, Using Tableau & Power BI for Database-Connected Interactive Dashboards, Advanced Features, Tableau LOD Expressions, Power BI DAX, Streaming Dashboards, Real-time stock market or IoT visualization, A/B Testing for Dashboard Efficacy, Iterative Design Insights.

Unit V - Data Storytelling, Communication, and Ethical Considerations - (12 Hours)

Introduction to Data Storytelling, Visual Communication and Design Principles, Narrative Structures and Presentation, Ethics in Data Science and AI, Responsible Communication and Social Impact.

Learning Resources

Text B

1. S. Few, Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press, 2012.
2. C. Ware, Information Visualization: Perception for Design. Morgan Kaufmann, 2012.
3. C. N. Knaflic, Storytelling with Data: A Data Visualization Guide for Business Professionals. Wiley, 2015.

Reference Books:

1. E. R. Tufte, The Visual Display of Quantitative Information. Graphics Press, 2001.
2. W. McKinney, Python for Data Analysis. O'Reilly Media.

SWAYAM / MOOC / YouTube Links

1. Tableau Public Gallery, [Online]. Available: https://public.tableau.com/gallery
2. Power BI Community, [Online]. Available: https://community.powerbi.com
3. FlowingData, [Online]. Available: https://flowingdata.com

4. Nightingale (Data Visualization Society Blog), [Online]. Available: https://nightingale.house
5. Coursera Data Visualization Courses, [Online]. Available: https://www.coursera.org/courses?query=d
6. edX Data Visualization Courses, [Online]. Available: https://www.edx.org/learn/data-visualization
7. Udacity Data Visualization Courses, [Online]. Available: https://www.udacity.com/courses/data-visualization

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-505-DS - Computational Laboratory - I		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 04 Hours/Week	02	Term Work : 25 Marks Oral : 25 Marks

Computational Laboratory - I (CL I) is companion course of theory core courses in Semester I. It is recommended that set of assignments or at least one mini-project/study project per course is to be completed. Set of problem statements are suggested. Course/ Laboratory instructors may frame suitable problem statements. Student has to submit a report/Journal consisting of appropriate documents - prologue, Certificate, table of contents, and other suitable write-up like (Introduction, motivation, aim and objectives, outcomes, brief theory, requirements analysis, design aspects, algorithms, mathematical model, complexity analysis, results, analysis and conclusions). Softcopy of report/journal and code is to be maintained by department/institute in digital repository.

Suitable platform/framework/language is to be used for completing mini- project/assignments.

Guidelines for Term Work Assessment

Continuous assessment of laboratory work is done based on performance of student. Each assignment/ mini-project assessment is to be done based on parameters with appropriate weightage. Suggested parameters for overall assessment as well as mini-project assessment include- timely completion, performance, innovation, efficient codes, usability, documentation and adhering to SDLC comprehensively.

Guidelines for practical Examination

It is recommended that practical examination should be conducted based on the understanding and knowledge of the subject as well as on the mini projects completed and the content understanding of laboratory work.

Suggested List of Laboratory Assignments

Mathematical Foundations

1. Data Handling and Descriptive Statistics:

- Load and explore real-world datasets (Iris, Titanic, etc.), Compute central tendencies (mean, median, mode), Compute dispersion (variance, standard deviation, IQR), Visualize distributions using histograms, boxplots, violin plots.

2. Hypothesis Testing

- Perform one-sample, two-sample t-tests, Perform Chi-square test for independence and goodness-of-fit, Apply ANOVA and post-hoc tests (Tukey's HSD), Use non-parametric tests (Mann-Whitney U, Wilcoxon signed-rank).

3. Correlation and Regression

- Compute Pearson and Spearman correlation, Simple and multiple linear regression, Assumptions checking: residual analysis, multicollinearity (VIF), Model selection using AIC, BIC.

4. Probability Distributions

- Simulate and visualize discrete (Binomial, Poisson) and continuous (Normal, Exponential) distributions, Calculate PDF, CDF, quantiles, Fitting a distribution to data and evaluating goodness-of-fit (Chi-square, KS-test)

5. Project Assignment

Choose a real-world dataset (healthcare, finance, social media, etc.)

- Perform end-to-end statistical analysis:
  - Data preparation
  - Exploratory analysis
  - Model fitting
  - Statistical inference
  - Presentation of findings using dashboards or Jupyter reports

Big Data Technologies

1. NoSQL Databases: HBase and MongoDB

- Install and configure HBase, Create tables, insert, retrieve and delete records, Connect HBase with Hive or Spark, Perform CRUD operations on MongoDB and query collections

2. Capstone Project

- Process a large real-world dataset (e.g., Twitter, weather, e-commerce logs), use HDFS + Hive + Spark to process and analyze the data, Store results in a NoSQL database and visualize outputs using Tableau/Power BI/Plotly

Machine Learning

1. Data Preprocessing, Supervised Learning & Model Evaluation:

A. Preprocessing:Load a real-world dataset (e.g., Titanic, UCI Heart Disease)

- Handle missing values, categorical encoding, and normalization
- Address class imbalance (e.g., SMOTE or class weighting)

B. Supervised Learning Models:

- Train Logistic Regression, k-NN, Decision Tree, and SVM
- Apply L1/L2 regularization and interpret results

C. Model Evaluation:

- Use accuracy, precision, recall, F1-score, ROC, AUC
- Perform train-test split, k-fold cross-validation

- Analyze overfitting and underfitting

D. Feature Selection:

- Apply filter methods (correlation), wrapper (RFE), and embedded (Lasso)

2. Clustering

- Apply k-Means, Hierarchical Clustering, and DBSCAN on datasets like Iris or Mall Customer Segmentation
- Evaluate using Silhouette Score, Davies-Bouldin Index, Adjusted Rand Index

3. Dimensionality Reduction:

- Apply PCA, t-SNE, and LDA on high-dimensional data (e.g., digit recognition)
- Visualize in 2D and interpret component contributions

4. Ensemble Learning & Hyperparameter Tuning

A. Bagging and Boosting:

- Implement Random Forest, AdaBoost, Gradient Boosting, and XGBoost
- Compare performance using cross-validation

B. Stacking:

- Build a stacked ensemble using base learners and meta-learner

C. Hyperparameter Optimization:

- Perform Grid Search, Random Search, and Bayesian Optimization using scikit-optimize.

Advanced Data Analysis and Visualization

1. Exploratory Data Analysis (EDA) in Python

- Use Pandas Profiling, seaborn, and matplotlib to generate visual summaries.

2. Interactive Dashboards with Plotly and Dash / Streamlit

- Build real-time or user-interactive analytics dashboards.

3. Data Storytelling with Tableau / Power BI

- Create storyboards and dynamic visuals with slicers, filters, and KPIs.

4. Geospatial Visualization

- Use Folium or Kepler.gl or Power BI to visualize location-based datasets like crime heatmaps or delivery routes.

5. Time Series Visualization

- Plot seasonal trends, moving averages, and anomalies in time series data (e.g., stock or energy data).

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-521A-DS - Optimization Techniques		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses : Foundation in data structures and algorithms, Programming, Understanding of discrete mathematics, probability, and linear algebra.

Course Objectives: The course aims to:

1. Develop concepts of optimization algorithms and iterative process leading to the solution.
2. Understand the optimization algorithms often used in data science.
3. Learn fundamentals and mathematics behind gradient based methods and to speed up the computation.
4. Develop idea of novel gradient free optimizers often used in AI/ML and data science.

Course Outcomes: Upon successful completion of this course, students will be able to:

- CO1: Differentiate between convex and non-convex functions
- CO2: Apply the gradient based methods for training neural network.
- CO3: Utilize advanced gradient based methods to perform faster computation.
- CO4: Ability to perform complex constrained optimization efficiently.

Course Contents

Unit I - BASICS OF OPTIMIZATION - (08 Hours)

Optimization Techniques, Optimal Problem Formulation, Convex Functions, First and Second Order Conditions for Optimizations, Convex and Non-Convex Optimization problems in Data Science.

Unit II - GRADIENT DESCENT TECHNIQUES - (08 Hours)

Mathematical background, Programming basic optimization problems and their solutions, Variants of Gradient Descent: Projected, Stochastic, Proximal, Accelerated, Coordinate Descent, Training a Neural Network.

Unit III - ADVANCED GRADIENT BASED METHODS - (08 Hours)

Advanced Gradient Descent: Projected, Stochastic, Proximal, Accelerated, Coordinate Descent, Training a Neural Network: Theory and Implementation, Newtons Method.

Unit IV - CONSTRAINED OPTIMIZATION - (08 Hours)

Kuhn-Tucker Conditions, Lagrangian Duality Theory, Penalty Function Method, Method of Multipliers, Random Search, Quadratic Programming

Unit V - GRADIENT FREE OPTIMIZATION TECHNIQUES - (08 Hours)

Genetic Algorithms: Operations, GA Operators, Real Coded-GA, Multi-objective GA, Particle Swarm Optimization, Evolutionary Strategy, Covariance Matrix Adaptation Methods, Differential Evolution Technique, Bayesian Optimization.

Learning Resources

Text Books

1. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.
2. K. Deb, Optimization for Engineering Design: Algorithms and Examples, 2nd ed., New Delhi, India: Prentice-Hall of India, 2012.
3. D. Roy and G. V. Rao, Stochastic Dynamics, Filtering and Optimization, Cambridge, U.K.: Cambridge University Press, 2017.
4. S. Sra, S. Nowozin, and S. J. Wright, Optimization for Machine Learning, Cambridge, MA, USA: MIT Press, 2012.

Reference Books:

1. R. K. Arora, Optimization: Algorithms and Applications, Boca Raton, FL, USA: CRC Press, 2015.
2. S. Bhandari, K. Sahay, and R. Singh, "Optimization Techniques in Modern Times and Their Applications," in Proc. IEEE Int. Conf. on Electrical, Electronics, Computers, Communication, Mechanical and Computing (IEECON), 2018, pp. 1–4, d
3. S. Sun, Z. Cao, H. Zhu, and J. Zhao, "A Survey of Optimization Methods From a Machine Learning Perspective," IEEE Trans. Cybern., vol. 50, no. 8, pp. 3668–3681, Aug. 2020,
4. H. A. Taha, Operations Research: An Introduction, 6th ed., New Delhi, India: Prentice-Hall of India, 1997.
5. R. W. Ott, Environmental Statistics and Data Analysis, Cambridge, U.K.: Cambridge University Press, 2020.

E-books:

1. "J. Xie, Optimization for Data Science, Lecture Notes, FS 23 [Online]. Available: https://n.ethz.ch/~jia
2. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

SWAYAM / MOOC / YouTube Links

1. https://onlinecourses.nptel.ac.in/noc23_cs64/preview

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-521B-DS - Advanced Information Retrieval		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses : Foundation in data structures and algorithms, Programming, Understanding of discrete mathematics, probability, and linear algebra.

Course Objectives: The course aims to:

1. To understand and apply classical and modern Information Retrieval models such as Boolean, Vector Space, probabilistic models, and neural retrieval techniques.
2. To design and implement Information Retrieval systems using indexing structures, embedding models, dense retrieval methods, and support for multilingual and domain-specific content.
3. To evaluate Information Retrieval systems using standard performance metrics such as precision, recall, MAP, and nDCG with benchmark datasets.
4. To build and deploy advanced applications such as semantic search engines and Retrieval-Augmented Generation systems using large language models and vector databases.

Course Outcomes: Upon successful completion of this course, students will be able to:

- CO1: Understand and apply classical and modern Information Retrieval models such as Boolean, Vector Space, probabilistic models, and neural retrieval techniques.
- CO2: Design and implement Information Retrieval systems incorporating indexing structures, semantic representations, neural embeddings, and support for multilingual and domain-specific data.
- CO3: Evaluate Information Retrieval systems using standard performance metrics and benchmark datasets, and analyze system behavior using user-centric evaluation methodologies.
- CO4: Develop and deploy advanced IR applications such as semantic search, conversational agents, and Retrieval-Augmented Generation systems using large language models and vector databases.

Course Contents

Unit I - Foundations of Information Retrieval and Classical Models - (08 Hours)

Introduction to Information Retrieval, IR problem definition, history, inverted index construction and Boolean query processing, skip pointers, Boolean model, Extended Boolean model, and Vector Space Model, TF, IDF, and TF-IDF, text preprocessing techniques -tokenization, stop word removal, lemmatization, stemming, and case folding, index structures: bi-word index and positional index, phrase queries, introduction to probabilistic retrieval and BM25.

Unit II - Web and Semantic Information Retrieval - (08 Hours)

Web IR fundamentals, structure of the web graph, crawler architecture, large-scale indexing, link analysis : PageRank and HITS, anchor text processing, near-duplicate detection using shingling and MinHash, structured data representation using metadata and schema.org, introduction to knowledge graphs, semantic search concepts, embedding-based representations using TF-IDF, Word2Vec, and SBERT, evaluation of web search quality.

Unit III - Evaluation and Benchmarks in Modern IR - (08 Hours)

IR evaluation principles and methodologies, precision, recall, F1-score, accuracy, mean average precision (MAP), mean reciprocal rank (MRR), and normalized discounted cumulative gain (nDCG), evaluation datasets and benchmarks: TREC, CLEF, FIRE, and BEIR, pseudo relevance feedback, query expansion techniques, user relevance modeling, A/B testing, inter-annotator agreement, human-in-the-loop evaluation strategies.

Unit IV - Neural, Dense, and Multilingual IR - (08 Hours)

Neural retrieval concepts: sparse to dense transition, embedding-based retrieval using Word2Vec, FastText, SBERT, and DPR, retrieval architectures: bi-encoders and cross-encoders ,ColBERT and SPLADE, multilingual IR, domain-specific IR in legal, medical, and rural governance applications, indexing with vector stores: FAISS.

Unit V - Retrieval-Augmented Generation and LLM-based IR - (08 Hours)

Retrieval-Augmented Generation architecture: chunking, embedding, vector store integration, and LLM-based generation, tools and framework: LangChain, LlamaIndex, OpenAI and HuggingFace pipelines, prompt engineering for IR-enhanced question answering, conversational search systems using memory-augmented agents, hybrid dense-sparse retrieval, semantic parsing for IR tasks, legal and governance chatbot applications.

Learning Resources

Text Books

1. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.
2. K. Deb, Optimization for Engineering Design: Algorithms and Examples, 2nd ed., New Delhi, India: Prentice-Hall of India, 2012.
3. D. Roy and G. V. Rao, Stochastic Dynamics, Filtering and Optimization, Cambridge, U.K.: Cambridge University Press, 2017.
4. S. Sra, S. Nowozin, and S. J. Wright, Optimization for Machine Learning, Cambridge, MA, USA: MIT Press, 2012.

Reference Books:

1. R. K. Arora, Optimization: Algorithms and Applications, Boca Raton, FL, USA: CRC Press, 2015.

2. S. Bhandari, K. Sahay, and R. Singh, “Optimization Techniques in Modern Times and Their Applications,” in Proc. IEEE Int. Conf. on Electrical, Electronics, Computers, Communication, Mechanical and Computing (IEECON), 2018, pp. 1–4, d

3. S. Sun, Z. Cao, H. Zhu, and J. Zhao, “A Survey of Optimization Methods From a Machine Learning Perspective,” IEEE Trans. Cybern., vol. 50, no. 8, pp. 3668–3681, Aug. 2020,

4. H. A. Taha, Operations Research: An Introduction, 6th ed., New Delhi, India: Prentice-Hall of India, 1997.

5. R. W. Ott, Environmental Statistics and Data Analysis, Cambridge, U.K.: Cambridge University Press, 2020.

E-books:

1. “J. Xie, Optimization for Data Science, Lecture Notes, FS 23 [Online]. Available: https://n.ethz.ch/~jia

2. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: https://www.deeplearningbook.org/

SWAYAM / MOOC / YouTube Links

1. https://onlinecourses.nptel.ac.in/noc23_cs64/preview

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-521C-DS - Quantum Computing		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses : Students should have prior knowledge of

1. Fundamentals of Programming (preferably Python)
2. Linear Algebra, Discrete Mathematics
3. Basic understanding of Artificial Intelligence and Machine Learning (for courses involving Quantum ML or hybrid algorithms)
4. Graph Theory (useful for optimization algorithms like QAOA)

Course Objectives: The course aims to:

1. To understand foundational concepts of the quantum technology.
2. To develop hands-on experience with experiments in quantum computing.
3. To explore quantum computational models and emerging applications for quantum computing.

Course Outcomes: After successful completion of this course, students will be able to:

- CO1: Comprehensive understanding of quantum computing concepts.
- CO2: Analyze the computation models.
- CO3: Model the circuits using quantum computation environments and frameworks.
- CO4: Understand the quantum operations such as noise and error-correction.
- CO5: Interpret the interfacing of classical and quantum technologies.

Course Contents

Unit I -Quantum Computation Fundamentals - (08 Hours)

Qubits, Superposition, Entanglement; Quantum Gates, Circuits, Measurement; Postulates of Quantum Mechanics; Tools: Qiskit or Cirq basics.

Unit II - Quantum Computing Architectures and Algorithms (08 Hours)

Guiding principles, Conditions for quantum computation, Harmonic oscillator quantum computer, Optical photon quantum computer, Optical cavity quantum electrodynamics, Ion traps, Nuclear magnetic resonance, Other implementation schemes, The quantum Fourier transform and its applications, Quantum search algorithms

Unit III - QUANTUM INFORMATION AND ERROR CORRECTION - (08 Hours)

Quantum noise and quantum operations: Classical noise and Markov processes, Quantum operations, Examples of quantum noise and quantum operations, Applications of quantum operations, Limitations of the quantum operations formalism, Introduction, The Shor code, Theory of quantum error-correction, Constructing quantum codes, Stabilizer codes, Fault-tolerant quantum computation.

Unit IV - ENTROPY AND INFORMATION THEORY - (08 Hours)

Entropy : Shannon Entropy, Basic properties of entropy, Von Neumann entropy, Strong sub additivity, Quantum information theory : Distinguishing quantum states and the accessible information, Data compression, Classical information over noisy quantum channels, quantum information over noisy quantum channels, Entanglement as a physical resource, Quantum cryptography

Unit V - Interfacing of Quantum and Classical Computing - (08 Hours)

Foundations of Quantum-Classical Interfaces, Variational Hybrid Algorithms, Programming hybrid systems, Challenges, Design Patterns, and Applications

Learning Resources

Text Books

1. M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information, 10th Anniversary ed. Cambridge, U.K.: Cambridge University Press, 2010.
2. C. Bernhardt, Quantum Computing for Everyone. Cambridge, MA, USA: MIT Press, 2019.
3. Y. Kitaev, A. Shen, and M. Vyalyi, Classical and Quantum Computation. Providence, RI, USA: American Mathematical Society, 2002.

Reference Books:

1. E. R. Johnston, N. Harrigan, and M. Gimeno-Segovia, Programming Quantum Computers. Sebastopol, CA, USA: O'Reilly Media, 2019.
2. IBM Q Team, Learn Quantum Computation Using Qiskit, 2020. [Online]. Available: https://qiskit.org/

SWAYAM / MOOC / YouTube Links

1. Hands-on quantum error correction with Google Quantum AI by Austin Fowler https://www.coursera.o error-correction?
2. Introduction to Quantum Computing: Quantum Algorithms and Qiskit | https://onlinecourses.nptel.ac

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-521D-DS - Computer Vision		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses : Fundamentals of Programming, Mathematics, Machine learning and image processing basics

Course Objectives: The course aims to:

1. To apply key image processing techniques.
2. To implement and evaluate feature detection
3. To develop object detection and matching algorithms
4. To design and evaluate end-to-end computer vision application

Course Outcomes: Upon successful completion of this course, students will be able to:

- CO1: Analyze visual data using image processing techniques.
- CO2: Extract and interpret features from images.
- CO3: Apply machine learning and deep learning models (e.g., CNNs) for tasks like image classification, detection, and segmentation.
- CO4: Develop end-to-end computer vision pipelines for real-world applications such as face recognition, surveillance, and autonomous driving.
- CO5: Evaluate and optimize vision systems using appropriate performance metrics and interpretability tools.

Course Contents

Unit I - Introduction to Computer Vision and Imaging Basics - (08 Hours)

Introduction to Computer Vision and Image Processing, Practical image data exploration and visualization, Basic image augmentation, Digital images, Pixels and color spaces, Image filtering and edge detection (Sobel, Canny), Image transformations, Libraries: OpenCV, PIL, skimage.

Unit II - Image Processing and Feature Extraction Techniques - (08 Hours)

Feature Detection and Description, Practical feature extraction for analytics pipelines, Harris and Shi-Tomasi corner detection, SIFT, SURF, ORB descriptors, Feature matching with brute-force and FLANN, Use cases in business analytics and simple object recognition.

Unit III - Machine Learning for Image Classification and Object Detection - (08 Hours)

Machine Learning for Vision, Classical ML models (k-NN, SVM, decision trees) for image classification, Dataset preparation and preprocessing, Introduction to CNNs (LeNet, AlexNet, VGG, ResNet), Transfer learning and fine-tuning, Tools: scikit-learn, TensorFlow/Keras basics.

Unit IV - Deep Learning in Computer Vision: CNNs and Architectures - (08 Hours)

Object Detection and Segmentation, Overview of object detection and segmentation, Sliding window and region proposals, Popular detectors (R-CNN, YOLO, SSD), Semantic and instance segmentation (FCNs, U-Net), Applications in retail analytics, healthcare image tagging.

Unit V - Real-World Applications and Case Studies in Computer Vision - (08 Hours)

Advanced Topics and Applications, Video analysis basics and motion detection, Face detection and recognition fundamentals, Ethics and privacy in computer vision, Introduction to 3D vision, Basic GAN concepts for data augmentation,

Learning Resources

Text Books

1. S. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed., Cham, Switzerland: Springer, 2022.

2. R. Szeliski, D. Forsyth, and J. Ponce, Computer Vision: A Modern Approach, 3rd ed., Pearson, 2023.

Reference Books:

1. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016

2. R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed., Hoboken, NJ, USA: Pearson, 2018

SWAYAM / MOOC / YouTube Links

1. Digital Image Processing, Prof. Prabir Kumar Biswas, IIT Kharagpur.
2. Computer Vision and Image Processing – Fundamentals and Applications - Course, by Prof. M. K. Bhuyan (IIT Guwahati)

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-521-DS - Skill Based Laboratory - I		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 02 Hours/Week	01	**Term Work** : 25 Marks **Oral** : 25 Marks

Skill Based Laboratory I is companion course of theory elective courses in Semester I. It is recommended that set of assignments and one capstone project per course is to be completed. Set of problem statements are to be suggested by the course teacher .

Course/ Laboratory instructors may frame suitable problem statements. Student has to submit a report/Journal consisting of appropriate documents- prologue, Certificate, table of contents, and other suitable write-up like (Introduction, motivation, aim and objectives, outcomes, brief theory, requirements analysis, design aspects, algorithms, mathematical model, complexity analysis, results, analysis and conclusions). Softcopy of report/journal and code is to be maintained by department/institute in digital repository.

Laboratory instructors may frame 7 to 10 assignments on the based of the elective course chosen by the student. Students are expected to design and develop an captone projects.

Savitribai Phule Pune University, Pune

Maharashtra, India

M. E. - Data Science (DS) - 2025 Pattern

Semester II

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-551-DS - Data Engineering		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses :

- Fundamentals of Programming Languages (Python) and databases (SQL)
- Data Structures, File systems
- Linux Command line usage

Course Objectives: The course aims to:

1. To equip students with the practical and theoretical foundations of building data pipelines.
2. To understand data modeling, ingestion, transformation, and orchestration.
3. To differentiate between batch and real-time data systems.
4. To explore tools for building scalable and efficient data solutions

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Explain the end-to-end data lifecycle and the roles in a modern data team.
CO2: Model and organize structured, semi-structured, and unstructured data for efficient querying.
CO3: Build robust data ingestion pipelines using tools like Kafka, NiFi, or custom APIs.
CO4: Process and transform large-scale data using frameworks like Apache Spark and Hive.
CO5: Orchestrate and monitor workflows using tools such as Apache Airflow and ensure data quality and reliability.

Course Contents

Unit I - Foundations of Data Engineering - (08 Hours)

Role of a data engineer; Data lifecycle from ingestion to consumption; Data quality: Accuracy, Completeness, Consistency, Validity, Uniqueness, Timeliness, Integrity, Data Profiling, Validation Rules, Anomaly Detection, Drift Detection, Schema Checks, Null Handling, Standardization; Stakeholders; Batch vs Streaming data; Data engineering vs Data science vs Data architecture, Data Cataloging, Data Storage Technologies : Architecture, Data center : Architecture, core elements, key characteristics, managing data centre.

Unit II - Data Modeling & Storage Systems - (08 Hours)

OLTP vs OLAP systems and use cases; Data lakes vs data warehouses vs lakehouses; Schema design: star, snowflake, denormalized, nested formats; Data partitioning, bucketing, indexing techniques for performance optimization; Storage systems: HDFS, NoSQL (MongoDB, Cassandra), Cloud Storage (AWS S3, GCP Cloud Storage, Azure Blob); Columnar vs row-based formats (Parquet, ORC, Avro); Compression and file format choices; Feature stores: architecture, online vs offline stores, use in ML workflows.

Case Study: Data lake for Healthcare data.

Unit III - Data Ingestion and Integration - (08 Hours)

ETL vs ELT; Data Ingestion Tools & Technologies: Apache NiFi, Sqoop, Kafka, Flume; Types of Data Ingestion: Structured and unstructured data ingestion; Data Ingestion Methods: REST APIs, Change Data Capture (CDC), streaming connectors.

Case Study: ETL/ELT pipelines for applications like Netflix/E-commerce platforms.

Unit IV - Data Processing & Transformation - (08 Hours)

Big Data Processing Frameworks: Apache Spark, Hive, and Presto; Data Wrangling: Data wrangling with Python/SQL; Job Scheduling & Workflow Orchestration: Directed Acyclic Graphs (DAGs), task dependencies, retries, triggers, and SLA monitoring; Performance tuning: partition pruning, cost-based optimization, indexing strategies.; Data versioning and lineage.

Unit V - Data Pipeline Orchestration and Monitoring - (08 Hours)

Workflow orchestration with Airflow and Luigi; DAGs, XComs, task dependencies, branching and conditional execution; Monitoring and logging tools: Prometheus, Grafana, ELK Stack, Stackdriver; Health checks, alerting and SLA tracking; Error handling and retries: exponential backoff, idempotency, checkpointing; Logging best practices; CI/CD for data workflows: Git, Docker, GitHub Actions, Jenkins, automated testing; Containerization of DAGs; Backfill and catch-up strategies; Version control and environment management; Parameterization and secrets management.

Learning Resources

Text Books

1. Reis, J., & Housley, M. (2022). Fundamentals of data engineering: Plan and build robust data systems. O'Reilly Media.
2. Kleppmann, M. (2017). Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. O'Reilly Media.
3. Akidau, T., Bradshaw, S., Chambers, C., Chernyak, S., Fernandez-Moctezuma, R. J., Lax, R., & Whittle, S. (2018). Streaming systems: The what, where, when, and how of large-scale data processing. O'Reilly Media.

Reference Books:

1. Crickard, P. (2020). Data engineering with Python: Work with massive datasets to design data models and automate data pipelines using Python. Packt Publishing.
2. Densmore, J. (2021). Data pipelines pocket reference: Moving and processing data for analytics. O'Reilly Media.
3. Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D & Dennison, D. (2015). Hidden technical debt in machine learning systems. In Advances in Neural Information Processing Systems (NeurIPS) . https://proceedings.neurips.cc/paper_files/paper/2015/file/86df7dcfd896fcaf26
Paper.pdf

SWAYAM / MOOC / YouTube Links

1. Google Cloud Data Engineering Certification
2. AWS Data Engineering tools documentation
3. Apache Airflow, Spark, and Kafka documentation.

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-552-DS - Cloud Computing for Data Analytics		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses : Students should have prior knowledge of

- Fundamentals of Programming, operating systems,
- Computer networks, data Analysis and machine learning workflows

Course Objectives: The course aims to:

1. To introduce the fundamentals of cloud computing and service models.
2. To understand virtualization, containers, and serverless computing.
3. To explore cloud-native data tools and deployment pipelines.
4. To apply cloud solutions in scalable AI/ML workflow

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Describe the architecture, models, and core services in cloud computing platforms.

CO2: Configure and deploy virtual machines and containers using platforms like Docker and Kubernetes.

CO3: Utilize cloud storage and compute services to handle scalable data workflows.

CO4: Design and implement data pipelines using cloud-native tools such as BigQuery, S3, and Lambda.

CO5: Integrate cloud-based ML services for developing, training, and deploying models at scale.

Course Contents

Unit I - Introduction to Cloud Computing - (08 Hours)

Definition and characteristics of cloud computing, Service Models: IaaS, PaaS, SaaS

Deployment Models: Public, Private, Hybrid, Community, Virtualization and Containers (Docker), Overview of Cloud Providers: AWS, Azure, GCP

Unit II - Cloud Infrastructure for Big Data and Analytics - (08 Hours)

Distributed computing concepts (Hadoop, Spark), Cloud storage systems: S3, HDFS, Azure Blob, Google Cloud Storage, Data Warehousing in the Cloud: Redshift, BigQuery, Snowflake, Compute Services: EC2, Lambda, Azure VMs, GCP Compute Engine, Real-time analytics architecture in cloud

Unit III - Data Analytics in the Cloud - (08 Hours)

Cloud-native analytics tools (AWS Glue, Azure Data Factory, GCP Dataflow), Batch vs Stream Processing: Apache Spark, Apache Beam, Data ingestion: Kinesis, Kafka, Pub/Sub, Analytics and Visualization: AWS QuickSight, Power BI, Google Data Studio, Data pipeline orchestration: Apache Airflow, Cloud Composer

Unit IV - Machine Learning on the Cloud - (08 Hours)

ML Services: AWS SageMaker, Azure ML Studio, Google AI Platform, Model deployment and serving, AutoML platforms, Monitoring ML models in cloud, Case Studies on Cloud-based ML pipelines

Unit V - Security, Governance & Cost Optimization in Cloud - (08 Hours)

Cloud security basics: IAM, encryption, policies, Data privacy, compliance (GDPR, HIPAA), Resource and cost management in cloud (Cost Explorer, Budgets), SLA, availability, disaster recovery, Best practices for cloud-based analytics applications

Learning Resources

Text Books

1. Rajkumar Buyya et al., Cloud Computing: Principles and Paradigms, Wiley
2. Srinivasan, Cloud Computing: A Hands-On Approach, Cambridge
3. Tom White, Hadoop: The Definitive Guide, O'Reilly
4. Learning Spark: Lightning-Fast Big Data Analysis, O'Reilly

Reference Books:

1. Zaharia, Matei et al., Databricks: Apache Spark and AI on the Cloud
2. Valliappa Lakshmanan, Sara Robinson, Michael Munn, Machine Learning Design Patterns - O'Reilly

SWAYAM / MOOC / YouTube Links

1. Cloud computing, By Prof. Soumya Kanti Ghosh | IIT Kharagpur Cloud computing - Course | https://nptel.ac.in/courses/106105167
2. Google Cloud Data Engineering https://www.coursera.org/professional-certificates/gcp-data-engineering

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-553-DS - Deep Learning		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite:

- Fundamentals of Programming Languages
- Mathematics

Course Objectives: The course aims to:

1. To introduce the fundamental mathematical concepts relevant to understand the deep learning.
2. To impart knowledge about the various deep learning approaches.
3. To introduce the working of various state of the art Deep Learning algorithms.
4. To learn to choose the appropriate Deep Learning algorithm to solve the given problem.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Construct a Neural Network for a Suitable application.
CO2: Selection of appropriate activation function for Deep Neural network.
CO3: Apply Techniques to improve neural network performance.
CO4: Understand functionality of all layers in a Convolutional Neural Network and Recurrent Neural Networks.

Course Contents

Unit I - Fundamentals of Neural Networks - (08 Hours)

Machine Learning Vs Deep Learning, Foundations of neural networks and deep learning, Perceptron and Multilayer Perceptron (MLPs), Activation function, Loss functions, Logistic regression as a neural network, different activation function, logistic regression cost function, logistic regression gradient descent, vectorizing logistic regression, forward and backward propagation.

Techniques to improve neural networks: regularization and optimizations, hyperparameter tuning, batch normalization, data augmentation, deep learning frameworks

Unit II - Deep Neural Networks and Regularization - (08 Hours)

Deep Neural Network Architectures, Weight Initialization: Zero Initialization, Random Initialization, Xavier Initialization (Glorot Initialization), He Initialization (Kaiming Initialization), LeCun Initialization, Orthogonal Initialization, Sparse Initialization, Batch Normalization, Hyperparameter Tuning, Dropout, Overfitting and Underfitting, Regularization Techniques: L1, L2, Early Stopping

Unit III - Convolutional Neural Networks - (08 Hours)

Convolutional Neural Networks & Different types, padding, dropout, strided convolution, pooling layers, convolutional implementation of sliding windows.

Architectures: LeNet, AlexNet, VGG, ResNet, Applications: Image Classification, Object Detection,

Unit IV - Recurrent Neural Networks (RNNs) and Sequence Models - (08 Hours)

Introduction to Sequential Data, Recurrent Neural Networks: Basic RNN, Bidirectional RNN, Vanishing Gradient Problem, Long Short-Term Memory (LSTM), Bi-directional LSTM and Gated Recurrent Units (GRU)

Applications: Time Series Forecasting, Sentiment Analysis

Unit V - Advanced Concepts in Deep Learning - (08 Hours)

Transfer Learning: Introduction to Transfer Learning, Transfer Learning Approaches, Popular Pretrained Models, Hands-on Implementation, Challenges and Best Practices for transfer learning,

Autoencoders (Denoising, Variational), Generative Adversarial Networks (GANs), Transformers: Transformer architecture, attention mechanism.

Deep Reinforcement Learning (DQN, Policy Gradients), Multimodal Learning (Vision-Language Models), Explainable AI (XAI) for Deep Learning, Ethical AI & Bias in Deep Learning

Learning Resources

Text Books

1. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
2. H. Hapke and C. Nelson, Building Machine Learning Pipelines: Automating Model Life Cycles with TensorFlow. Sebastopol, CA, USA: O'Reilly Media, 2020.

Reference Books:

1. 1. C. Chen, N. R. Murphy, K. Parisa, D. Sculley, and T. Underwood, Reliable Machine Learning: Applying SRE Principles to ML in Production. Sebastopol, CA, USA: O'Reilly Media, 2022.

SWAYAM / MOOC / YouTube Links

1. Deep Learning By Prof. Prabir Kumar Biswas https://onlinecourses.nptel.ac.in/noc20_cs62/preview

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-554-DS - Computational Laboratory - II		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 04 Hours/Week	02	**Term Work** : 25 Marks **Oral** : 25 Marks

Computational Laboratory II (CL II) is companion course of theory core courses in Semester II. It is recommended that set of assignments or at least one mini-project/study project per course is to be completed. Set of problem statements are suggested. Course/ Laboratory instructors may frame suitable problem statements. Student has to submit a report/Journal consisting of appropriate documents - prologue, Certificate, table of contents, and other suitable write-up like (Introduction, motivation, aim and objectives, outcomes, brief theory, requirements analysis, design aspects, algorithms, mathematical model, complexity analysis, results, analysis and conclusions). Softcopy of report/journal and code is to be maintained by department/institute in digital repository.

Suitable platform/framework/language is to be used for completing mini-project/assignments.

Guidelines for Term Work Assessment

Continuous assessment of laboratory work is done based on performance of student. Each assignment/ mini-project assessment is to be done based on parameters with appropriate weightage. Suggested parameters for overall assessment as well as mini-project assessment include- timely completion, performance, innovation, efficient codes, usability, documentation and adhering to SDLC comprehensively.

Guidelines for practical Examination

It is recommended that practical examination should be conducted based on the understanding and knowledge of the subject as well as on the mini projects completed and the content understanding of laboratory work.

Suggested List of Laboratory Assignments

Data Engineering

1. **Data Ingestion Pipeline**
   - Use Apache NiFi or custom Python script to pull data from APIs and databases into a data lake (S3/HDFS).
2. **ETL Pipeline using Apache Spark**
   - Load, clean, transform and store structured/unstructured data using PySpark.
3. **Streaming Data Processing with Kafka and Spark Streaming**
   - Simulate a real-time data stream (e.g., IoT sensor or stock ticker) and process using Spark.
4. **Data Warehousing with Snowflake/Redshift/BigQuery**
   - Load data into a cloud data warehouse, design star/snowflake schema, and perform analytical queries.

5. **Workflow Orchestration using Apache Airflow**

- Design a DAG that automates a data pipeline including ingestion, processing, and logging.

Deep Learning

1. Design and implement logic gates (AND, OR, NOR, XOR) using a Perceptron model.

- Design and implement basic logic gates (AND, OR, NOR, XOR) using a Perceptron model. Understand how single-layer and multi-layer perceptrons can learn simple logical operations and explore the limitations of linear classifiers.

2. To observe how different weight initialization strategies affect the training behavior of a neural network using a simple 2D classification task.

- Use the make_moons dataset from sklearn.datasets, which generates a non-linearly separable two-class dataset ideal for visualizing decision boundaries and convergence behavior.

3. MNIST Digit Classification with a Feedforward NN

- Build a neural network using PyTorch/TensorFlow to classify handwritten digits (MNIST dataset).

4. Text Generation with LSTM

- Train an LSTM to generate Shakespeare-like text.

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PCC-561A- DS - Generative AI		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Course : Students should have prior knowledge of Fundamentals of NLP, Deep learning, Programming

Course Objectives: The course aims to:

1. To define and describe fundamental Generative modeling techniques (GANs, VAEs, Transformers, Diffusion Models)
2. To implement and fine-tune Generative models using modern ML frameworks.
3. To analyze model performance and behavior using quantitative and qualitative evaluation techniques.
4. To evaluate the ethical, legal, and societal implications of generative AI applications.
5. To design and create end-to-end generative AI solutions for real-world use cases

Course Outcomes: After successful completion of this course, students will be able to:
CO1: Develop mathematical foundations for generative AI
CO2: Implement and fine-tune generative models using Hugging Face, langChain.
CO3: Comprehend the issues in the existing architectures of generative AI.
CO4: Understand the RAG implementation for contextual responses.
CO5: Apply generative AI techniques to tasks such as image synthesis, text generation, and data augmentation.

Course Contents

Unit I - Foundations of Generative AI - (07 Hours)

Discriminative vs. Generative models, Latent variable models, Maximum Likelihood Estimation (MLE), KL divergence, ELBO, Basic probabilistic models: Gaussian Mixture Models, Hidden Markov Models
Case Study: Implement a basic Gaussian Mixture Model and sample from it

Unit II - Probabilistic Models & VAEs - (07 Hours)

GAN architecture: Generator and Discriminator, Loss functions, minimax optimization, Variants: DCGAN, Conditional GANs, StyleGAN, CycleGAN, Challenges: Mode collapse, training instability, Lab: Implement a DCGAN for MNIST or CIFAR-10
Case study: Image-to-image translation using CycleGAN

Unit III - Variational Autoencoders (VAEs) and Diffusion Models - (07 Hours)

Autoencoders and limitations, Introduction to VAEs: Probabilistic encoding, reparameterization trick, Denoising Diffusion Probabilistic Models (DDPM), Latent Diffusion, Stable Diffusion overview

Lab: Train a VAE on Fashion-MNIST

Project Idea: Text-to-image generation using latent diffusion models

Unit IV - Generative Transformers and Large Language Models - (07 Hours)

Language modeling and autoregressive generation, Transformer architecture (Self-attention, Positional encoding),GPT family, T5, BERT vs GPT, Fine-tuning, prompt engineering, in-context learning,

Lab: Use HuggingFace Transformers to fine-tune a GPT-2 model

Project Idea: Story or poetry generation using GPT-2 or GPT-3 API

Unit V - Applications, Safety, and Ethics in Generative AI - (07 Hours)

Applications: Art, code (Codex), audio (Jukebox), video, medicine, Evaluation metrics: Inception Score, FID, BLEU, ROUGE, Risks: Misinformation, bias, deepfakes, copyright, Guardrails: RLHF, Vector databases, RAG, watermarking, policy and governance frameworks

Lab: Build a fake image detector using CNNs

Discussion: Case studies on AI hallucination and model misuse

Learning Resources

Text Books

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
2. Foster, D. (2022). Generative deep learning: Teaching machines to paint, write, compose, and play (2nd ed.). O'Reilly Media.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008). https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee9
Paper.pdf

Reference Books:

1. Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114.
https://arxiv.org/abs/1312.6114
2. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. arXiv preprint arXiv:2006.11239. https://arxiv.org/abs/2006.11239
3. OpenAI. OpenAI API documentation. Retrieved May 21, 2025, from https://platform.openai.com/docs
4. Hugging Face. Hugging Face documentation. Retrieved May 21, 2025, from https://huggingface.co/do

SWAYAM / MOOC / YouTube Links

1. Generative AI with Large Language Models - https://www.coursera.org/learn/generative-ai-with-llms
2. Programming with Generative AI – NPTEL (IISc Bangalore)
3. Generative AI and Large Language Models – SWAYAM

4. Generative AI: Prompt Engineering Basics - https://www.coursera.org/learn/generative-ai-prompt-engineering-for-everyone

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-561B-DS - Blockchain Technology		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses :

- Linear Algebra
- Basic Probability and Statistics
- Foundations of Machine Learning

Course Objectives: The course aims to:

1. To understand the core concepts, architecture, and functioning of blockchain technology.
2. To explore consensus algorithms and their role in maintaining distributed ledgers.
3. To study smart contracts and decentralized applications (D-Apps).
4. To analyze the integration of blockchain in data systems for trust, transparency, and auditability.
5. To examine recent advancements in data storage, processing, and sharing mechanisms.
6. To gain practical experience using blockchain platforms and next-generation data systems.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Understand foundational concepts of blockchain architecture and cryptographic tools.
CO2: Develop and deploy smart contracts using Ethereum/Solidity and Hyperledger.
CO3: Analyze blockchain-based solutions in real-world domains like healthcare and DeFi.
CO4: Explore decentralized file systems and blockchain-based data sharing.
CO5: Integrate blockchain with modern data platforms and architectures.

Course Contents

Unit I - Introduction to Blockchain - (07 Hours)

Basics of Blockchain: Blocks, Hashes, Chains, Transactions, Public vs Private Blockchains, Cryptographic foundations: Hashing, Public-Key Cryptography, Merkle Trees, Blockchain structure and data immutability, Distributed Ledger Technologies (DLT): basics

Unit II - Distributed Ledger Technologies (DLT) - (07 Hours)

Architecture of DLTs: Nodes, Transactions, and Blocks, Popular Platforms: Bitcoin, Ethereum, Hyperledger Fabric, Peer-to-Peer Networks and IPFS, Challenges in Decentralized Data Sharing

Unit III - Smart Contracts and Blockchain Applications - (07 Hours)

Smart Contracts: Design, Deployment, and Testing, Ethereum and Solidity programming, Hyperledger Fabric and chaincode, Use Cases: Voting systems, DeFi, Healthcare, Supply Chain, Token Standards: ERC-20, ERC-721 (NFTs)

Unit IV - Blockchain with Modern Data Architectures - (07 Hours)

Modern Data Architecture: Data Lakehouse, Data Mesh, Delta Lake, Apache Hudi, Apache Iceberg, Real-time streaming systems: Apache Kafka, Apache Pulsar, Object Storage Systems: AWS S3, MinIO, Data Governance and Provenance

Unit V - Blockchain for Data Integrity and Decentralized Storage - (07 Hours)

Blockchain for Data Integrity, Provenance, and Sharing, InterPlanetary File System (IPFS): hands-on and applications, Decentralized Storage Systems (Filecoin, Arweave), Case Studies: Blockchain in healthcare data, audit trails, and finance, Project Work: Build a prototype integrating blockchain with a modern data pipeline

Learning Resources

Text Books

1. Imran Bashir, Mastering Blockchain.
2. Daniel Drescher, Blockchain Basics.
3. Stephen Grider, Ethereum and Solidity: The Complete Developer's Guide.
4. Martin Kleppmann, Designing Data-Intensive Applications.
5. Hyperledger & Ethereum Developer Documentation, [Online].
6. Apache Iceberg, Delta Lake, and Kafka Documentation, [Online].

Reference Books:

1. Andreas M. Antonopoulos, Mastering Bitcoin: Unlocking Digital Cryptocurrencies, O'Reilly Media, 2017.
2. Roger Wattenhofer, The Science of the Blockchain, CreateSpace Independent Publishing Platform, 2016.
3. Melanie Swan, Blockchain: Blueprint for a New Economy, O'Reilly Media, 2015

SWAYAM / MOOC / YouTube Links

1. Blockchain Technology — IIT Madras -https://onlinecourses.nptel.ac.in/noc22_cs93/preview
2. Blockchain and Cryptocurrency — IIT Kanpur -https://onlinecourses.nptel.ac.in/noc21_cs68/preview
3. Introduction to Cryptography and Security — IIT Kanpur -https://onlinecourses.nptel.ac.in/noc21_cs61/p
4. Decentralized Finance and Blockchain Technology — IIT Madras (periodic/offered sometimes)

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-561C-DS Web Mining and Analytics		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses :

· Basic programming,

· Data structure

· Statistics

Course Objectives: The course aims to:

1. To comprehend the fundamental concepts, architectures, and algorithms underpinning web mining and analytics.
2. To apply various web data collection, pre-processing, and analysis techniques for extracting valuable insights from web content, structure, and usage.
3. To implement practical web mining algorithms and analytical methods using appropriate tools and programming languages.
4. To design and develop solutions for real-world web-related problems, including recommendation systems, sentiment analysis, and search engine optimization.
5. To evaluate the performance and ethical implications of web mining algorithms and analytical models.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Analyze and preprocess diverse web data sources (content, structure, usage) for effective mining and analytical tasks.

CO2: Apply appropriate web mining algorithms to extract patterns, trends, and knowledge from web data, including text, hyperlinks, and user behavior.

CO3: Develop and implement practical web analytics solutions for tasks such as sentiment analysis, user profiling, and recommendation systems.

CO4: Evaluate the performance of web mining and analytics models, addressing scalability, privacy, and ethical considerations.

CO5: Implement modern AI techniques such as deep learning and transformer models in web analytics applications.

Course Contents

Unit I - Introduction to Web Mining and Analytics - (07 Hours)

Introduction, Definition, Motivation, Challenges, Applications, Web Mining vs. Data Mining, Web Mining Types, Content, Structure, Usage, Web Analytics Overview, Key Metrics, Web Analytics Process, Web Architecture Overview, HTTP, HTML, XML, Web Services, APIs, Web Data Collection, Web logs, clickstream data, user profiles, social media data, web scraping, crawler architectures, Ethical

Web Mining, Laws, Privacy, Governance, Introduction to real-time web analytics and streaming web data.

Unit II - Web Content Mining - (07 Hours)

Introduction to Web Content Mining, Challenges, Text Preprocessing: Tokenization, Stop Word Removal, Stemming, Lemmatization, Part-of-Speech Tagging, Feature Engineering for Text Data: TF-IDF, Word Embeddings (Word2Vec, GloVe), Text Classification: Naive Bayes, SVMs, Neural Networks for Text, Web Document Clustering: K-Means, Hierarchical Clustering, Topic Modelling: LSA, LDA, Information Extraction from Web Pages: Named Entity Recognition (NER), Relation Extraction, Wrapper Induction, Sentiment Analysis and Opinion Mining: Lexicon-based, Machine Learning-based Approaches, Deep Learning for Sentiment Analysis, Transformer-based models (BERT).

Unit III - Web Structure Mining and Link Analysis - (07 Hours)

Introducing Web Structure Mining, Web Graph Models: Directed, Undirected, Bipartite Graphs, Algorithms for Link Analysis: Web graph community detection, PageRank, HITS, SALSA, Link Spam Detection: Girvan-Newman Algorithm, Link-based Classification and Clustering, Modularity Maximization, Web document organization using link analysis, search engine ranking based on link analysis, Fundamentals of Social Network Analysis, Centrality Measures, Influence Propagation, Graph Neural Networks for Link Prediction, Advanced Influence Detection Models.

Unit IV - Web Usage Mining and User Behaviour Analysis - (07 Hours)

Introduction to Web Usage Mining, Web Log Data Formats, Data Collection and Preprocessing: Data Cleaning, Session Identification, Path Completion, User Identification, Web Usage Association Rule Mining: Apriori, FP-Growth, Sequential Pattern Mining: GSP, PrefixSpan, Web User Clustering: K-Means, DBSCAN, Collaborative Filtering for Recommendation Systems: User-Based, Item-Based, Matrix Factorization (SVD), Personalization and Adaptive Web Sites, User Profiling and Segmentation, Clickstream Analysis, Conversion Funnel Analysis, Reinforcement Learning for Personalized Recommendations.

Unit V - Web Analytics, Trends, and Applications - (07 Hours)

Web Analytics Tools, Google Analytics, Adobe Analytics Overview, Page Views, Unique Visitors, Bounce Rate, Conversion Rate, Exit Rate, Time on Site, A/B Testing and Multivariate Testing for Web Optimization, Search Engine Optimization (SEO) Principles, Keyword Analysis, On-page and Off-page SEO, Web Advertising and Click Fraud Detection, Key Web Analytics Metrics, Introduction to Social Media Mining, Influencer Identification, Trend Detection, Web Mining and Analytics Ethics, Privacy, Discrimination, Data Ownership, Web Mining Trends, Deep Learning for Web Data, Real-time Web Analytics, Semantic Web Technologies, Ethical Considerations in SEO Manipulation and Ad Fraud Detection.

Learning Resources

Text B

1. Liu, B. (2011). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer.
2. Aggarwal, C. C. (2011). Social Network Data Analytics. Springer.

Reference Books:

1. Manning, C. D., Raghavan, P., & Schütze, H. (2007). Introduction to Information Retrieval. Cambridge University Press.

2. Russell, M. A. (2014). Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites. O'Reilly Media.

SWAYAM / MOOC / YouTube Links

1. Google Analytics official guides and tutorials, https://analytics.google.com/analytics/academy/

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-561D-DS - Data Privacy and Ethics		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses :

1. Computer Networks
2. Programming
3. Data Structures

Course Objectives: The course aims to:

1. To understand the fundamental concepts of data privacy, security, and the associated ethical challenges in modern computing systems.
2. To comprehend various data privacy regulations and frameworks, including their technical, legal, and social implications.
3. To apply privacy-enhancing technologies (PETs) and data anonymization techniques to design privacy-preserving data management and analysis solutions.
4. To analyze and mitigate privacy risks and vulnerabilities in data collection, processing, and sharing across different domains.
5. To critically evaluate the ethical dimensions of emerging technologies and data practices, fostering responsible innovation.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Identify and analyze privacy risks and vulnerabilities in data-intensive systems, considering various attack models and potential harms.

CO2: Apply fundamental privacy-preserving techniques, including cryptographic methods and anonymization algorithms, to secure sensitive data.

CO3: Evaluate and ensure compliance with major data privacy regulations (e.g., GDPR, HIPAA, CCPA) in system design and data handling.

CO4: Develop and critically assess ethical frameworks for data collection, usage, and algorithmic decision-making in diverse real-world scenarios.

CO5: Implement privacy-aware AI models, addressing algorithmic bias, transparency, and fairness in decision-making systems

Course Contents

Unit I - Fundamentals of Data Privacy and Ethical Foundations - (07 Hours)

Introduction to Data Privacy and types, Privacy vs Security vs Confidentiality, Evolution of privacy concerns, Major data breaches and lessons learned, Basics of computing ethics and ethical frameworks, Professional ethics, Human rights and privacy, Case studies including GDPR and AI surveillance ethics.

Unit II - Data Privacy Regulations and Legal Frameworks - (07 Hours)

Overview of global privacy laws: GDPR, CCPA/CPRA, HIPAA, data subject rights and business obligations, Consent and compliance mechanisms, Cross-border data transfer rules, Emerging AI-related privacy regulations.

Unit III - Privacy-Enhancing Technologies (PETs) and Anonymization - (07 Hours)

Core PETs: Data minimization, anonymization vs pseudonymization, k-anonymity, differential privacy basics, Privacy-preserving cryptographic techniques (homomorphic encryption, secure multi-party computation), Federated learning, Data masking and synthetic data, Challenges and risks.

Unit IV - Privacy in Data Management and Applications - (07 Hours)

Database and cloud privacy controls, Access control models, Privacy in big data and IoT, Data collection and consent management, Social media and mobile privacy, Tracking technologies and user consent platforms, Decentralized identity with blockchain, AI-driven privacy risk assessments.

Unit V - Ethical AI and Algorithmic Bias - (07 Hours)

Ethics in AI: fairness, accountability, transparency, explainability (FATE), Sources and mitigation of algorithmic bias, Explainable AI techniques, Ethical concerns in surveillance, Marketing, healthcare data, Digital identity privacy, Roles of data ethicists and privacy engineers.

Learning Resources

Text Books

1. D. J. Solove and P. M. Schwartz, Privacy Law Fundamentals, 3rd ed. IAPP, 2020.
2. O'Reilly, R., & Schmidt, D. C. Privacy-Preserving Data Publishing: An Overview. Morgan & Claypool Publishers. 2017
3. Dwork, C., & Roth, A. The Algorithmic Foundations of Differential Privacy. Now Publishers Inc.2014

Reference Books:

1. Koops, B.-J. (2011). Privacy and the Internet. Edward Elgar Publishing.
2. Cavoukian, A., & Castro, D. J. (2011). Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada.
3. European Data Protection Board (EDPB). (n.d.). General Data Protection Regulation (GDPR) official documentation and guidelines. Retrieved [Date], from https://edpb.europa.eu/

SWAYAM / MOOC / YouTube Links

1. Data Privacy and Security-https://nptel.ac.in/courses/106/105/106105215/
2. Information Security and Privacy -https://nptel.ac.in/courses/106/104/106104191/

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-562A-DS - Financial Analytics		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Course Objectives: The course aims to:

1. To provide students with a solid foundation in financial systems, markets, and instruments.
2. To enable learners to acquire skills in financial data preprocessing and time-series modeling.
3. To train students to apply machine learning techniques to solve real-world financial problems.
4. To introduce algorithmic trading, portfolio optimization, and emerging trends in financial analytics.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Analyze and interpret various financial datasets for informed decision-making.
CO2: Apply time-series forecasting and ML techniques to financial use cases
CO3: Build predictive models for credit risk, fraud detection, and stock price prediction.
CO4: Develop financial strategies using algorithmic trading and evaluate them through backtesting.

Course Contents

Unit I - Fundamentals of Financial Systems and Markets - (07 Hours)

Overview of financial markets and instruments, Introduction to financial statements and financial ratios, Risk and return concepts, Time value of money and discounted cash flows, Basics of portfolio theory, Regulatory framework and financial compliance (SEBI, Basel norms)

Unit II - Data Sources and Preprocessing in Finance - (07 Hours)

Types and sources of financial data (stock prices, fundamentals, alternative data), Data cleaning and transformation for financial datasets, Handling time-series data and missing values, Feature engineering for financial variables, Financial metrics: beta, alpha, Sharpe ratio, volatility, etc.

Unit III - Time Series Forecasting and Modeling - (07 Hours)

Stationarity, ACF/PACF, ARIMA, SARIMA models, Exponential smoothing models (Holt, Holt-Winters), GARCH models for volatility prediction, Application of LSTM and RNN for financial forecasting, Evaluation metrics for time-series models (MAPE, RMSE, etc.)

Unit IV - Machine Learning for Financial Analytics - (07 Hours)

Credit scoring using classification models (Logistic Regression, Random Forest, XGBoost), Fraud detection using anomaly detection techniques, Clustering for customer segmentation (K-means, DBSCAN), Predicting stock prices using regression and ensemble methods, Model validation, overfitting, cross-validation in finance

Unit V - Advanced Topics in Financial Analytics - (07 Hours)

Algorithmic trading strategies and backtesting, Portfolio optimization using ML and risk management techniques, Sentiment analysis on financial news and social media, Introduction to Quantitative Finance libraries (QuantLib, Pyfolio, Backtrader), Ethics and limitations of AI/ML in financial decision making

Learning Resources

Text B

1. “Financial Analytics with R” – Mark J. Bennett, Dirk L. Hugen
2. “Machine Learning for Asset Managers” – Marcos Lopez de Prado
3. “Financial Risk Modelling and Portfolio Optimization with R” – Bernhard Pfaff

Reference Books:

1. “Quantitative Financial Analytics: The Path to Investment Profits” – Kenneth L. Grant
2. “Advances in Financial Machine Learning” – Marcos Lopez de Prado
3. “Python for Finance: Mastering Data-Driven Finance” – Yves Hilpisch
4. “Statistics and Data Analysis for Financial Engineering” – David Ruppert

SWAYAM / MOOC / YouTube Links

1. Machine Learning for Finance" by Prof. P.K. Sinha https://nptel.ac.in/courses/110/104/110104124
2. Financial Institutions and Markets" by Prof. S.S. Srikant https://nptel.ac.in/courses/110/106/1101060

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-562B-DS - Healthcare Analytics		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses :Graph Theory, Probability & Statistics, Programming, Data Mining

Course Objectives: The course aims to:

1. To provide a foundational understanding of healthcare systems and the types of data they generate.
2. To enable students to perform effective data preprocessing and exploratory data analysis on healthcare datasets.
3. To apply machine learning models for disease prediction and clinical decision support.
4. To develop awareness of ethical, privacy, and legal issues surrounding healthcare analytics.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Understand and distinguish between various healthcare data types and structures.

CO2: Apply data preprocessing techniques and generate insights from real-world health data.

CO3: Build and evaluate predictive models for healthcare applications using machine learning.

CO4: Critically analyze ethical challenges and ensure compliance with legal frameworks in healthcare data handling.

Course Contents

Unit I - Introduction to Healthcare Systems and Data - (07 Hours)

Overview of healthcare systems (global and Indian context), Types of healthcare data: clinical, administrative, claims, patient-generated, Electronic Health Records (EHR), HL7, ICD codes, Healthcare data sources: hospitals, public health, wearables, biosensors, Challenges in healthcare data: privacy, interoperability, sparsity

Unit II - Data Preprocessing and Exploratory Analysis - (07 Hours)

Data cleaning techniques for healthcare datasets, Handling missing values, noise, and outliers, Exploratory Data Analysis (EDA) in healthcare, Feature engineering for medical data, Visualization techniques for clinical insights.

Unit III - Predictive Analytics in Healthcare - (07 Hours)

Disease prediction models (diabetes, cancer, heart disease), Classification and regression techniques in healthcare, Time-series analysis of patient data, Evaluation metrics: ROC, AUC, sensitivity, specificity, Model interpretability and explainability (SHAP, LIME)

Unit IV - Machine Learning for Clinical Decision Support - (07 Hours)

ML applications in diagnostics and prognosis, Risk scoring and early warning systems, NLP in healthcare: clinical notes, patient feedback, Recommender systems for personalized treatment, Deep learning in medical imaging (X-ray, MRI, CT scans)

Unit V - Ethical, Legal, and Emerging Trends in Healthcare Analytics - (07 Hours)

Data privacy laws (HIPAA, GDPR, DISHA in India), Ethical issues in healthcare analytics, Bias, fairness, and transparency in AI-driven diagnosis, Remote patient monitoring, Telehealth, mHealth, AI in drug discovery, genomics, and pandemic response

Learning Resources

Text Books

1. Healthcare Analytics: From Data to Knowledge to Healthcare Improvement", Hui Yang, Eva K. Lee, Wiley
2. "Big Data Analytics in Healthcare", Katherine Marconi, Harold Lehmann, Jones & Bartlett Learning
3. "Machine Learning for Healthcare" Ramesh Natarajan, Parashar Shah, Packt Publishing

Reference Books:

1. "Artificial Intelligence in Healthcare", Adam Bohr, Kaveh Memarzadeh, Academic Press
2. "Predictive Analytics in Healthcare: A Guide to Data Science for Healthcare Professionals", Jason Burke, Wiley
3. "Biomedical Informatics: Computer Applications in Health Care and Biomedicine", Edward H. Shortliffe, James J. Cimin, Springer

SWAYAM / MOOC / YouTube Links

1. Introduction to Biomedical Engineering https://nptel.ac.in/courses/102103044
2. AI for Medicine Specialization – https://www.coursera.org/specializations/ai-for-medicine

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-562C-DS - Social Network Analytics		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses :Graph Theory, Probability & Statistics, Programming, Data Mining

Course Objectives: The course aims to:

1. Understand the fundamental concepts of social networks and network theory.
2. Model and analyze real-world social networks using graph-theoretic and statistical approaches.
3. Apply centrality, community detection, and influence maximization methods.
4. Use modern tools and techniques to visualize, interpret, and extract insights from social network data.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Understand the structure and representation of social networks using graphs and matrices.
CO2: Analyze social networks using metrics like centrality, clustering, and shortest paths.
CO3: Apply algorithms for community detection, influence propagation, and link prediction.
CO4: Use tools and frameworks for social network analysis and apply them to real-world problems.

Course Contents

Unit I - Introduction to Social Networks - (07 Hours)

Introduction, Definition, Motivation, Challenges, Applications, Web Mining vs. Data Mining, Web Mining Types, Content, Structure, Usage, Web Analytics Overview, Key Metrics, Web Analytics Process, Web Architecture Overview, HTTP, HTML, XML, Web Services, APIs, Web Data Collection, Web logs, clickstream data, user profiles, social media data, web scraping, crawler architectures, Ethical Web Mining, Laws, Privacy, Governance, Introduction to real-time web analytics and streaming web data.

Unit II - Network Metrics and Properties - (07 Hours)

Degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, density, diameter, clustering coefficient, average path length, assortativity, homophily and influence.

Unit III - : Community Detection and Graph Partitioning - (07 Hours)

Cohesive subgroups, cliques and k-cores, modularity, Girvan-Newman algorithm, Louvain method, spectral clustering, hierarchical clustering, graph partitioning techniques, evaluation of communities.

Unit IV - Information Diffusion and Influence Models - (07 Hours)

Diffusion models: SI, SIR, SIS, threshold and cascade models, influence maximization, independent cascade model, linear threshold model, applications in viral marketing and opinion dynamics.

Unit V - Applications and Tools for Social Network Analysis - (07 Hours)

Social network visualization and analysis tools including Gephi, NetworkX, SNAP, case studies in Twitter, Facebook, citation and collaboration networks, ethical considerations and privacy in social data, future trends and research directions in social network analysis.

Learning Resources

Text Books

1. Easley, D. & Kleinberg, J., Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press
2. Newman, M. E. J., Networks: An Introduction, Oxford University Press
3. Wasserman, S. & Faust, K., Social Network Analysis: Methods and Applications, Cambridge University Press

E-References:

1. Gephi: https://gephi.org
2. NetworkX: https://networkx.github.io
3. SNAP: https://snap.stanford.edu
4. UCINET: https://sites.google.com/site/ucinetsoftware/home
5. Social Network Analysis tutorials: https://towardsdatascience.com (search: social network analysis)

SWAYAM / MOOC / YouTube Links

1. Social Networks – by Prof. Sandeep Sen, IIT Delhi: https://nptel.ac.in/courses/106/102/106102132
2. Introduction to Graph Theory – by Prof. Sudarshan Iyengar, IIT Ropar: https://nptel.ac.in/courses/106
3. Data Science for Engineers – by Prof. Raghunathan Rengaswamy, IIT Madras: https://nptel.ac.in/cours

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
PEC-562D-DS - IoT Analytics		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 03 Hours/Week	03	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite Courses : Data Analytics, IoT Fundamentals, Programming, Statistics

Course Objectives: The course aims to:

1. Understand the role of data analytics in the Internet of Things ecosystem.
2. Apply statistical and machine learning methods for analyzing IoT data.
3. Develop analytics solutions for real-time, streaming, and sensor data.
4. Utilize tools and platforms for processing and visualizing IoT-generated data.

Course Outcomes: Upon successful completion of this course, students will be able to:

CO1: Understand IoT data characteristics and processing needs for various applications.

CO2: Apply statistical and machine learning techniques for descriptive and predictive analytics on IoT data.

CO3: Develop and deploy real-time analytics solutions using stream and edge computing frameworks.

CO4: Use tools and platforms for storing, processing, and visualizing IoT analytics workloads.

Course Contents

Unit I - IoT Data Characteristics and Sources - (07 Hours)

IoT system architecture and components, sources of IoT data including sensors, devices, and gateways, types of IoT data: structured, semi-structured, unstructured, temporal and spatial data, data sampling, aggregation, and quality challenges, data acquisition and transmission protocols.

Unit II - Descriptive Analytics for IoT - (07 Hours)

Data preprocessing and cleaning, exploratory data analysis techniques, summary statistics and visualization, correlation and anomaly detection, time-series decomposition, trend analysis, seasonality in sensor data.

Unit III - Predictive Analytics for IoT - (07 Hours)

Machine learning for IoT: regression, classification, clustering, decision trees, random forests, SVMs, k-means, use of deep learning methods for time-series forecasting, model evaluation techniques, case studies in predictive maintenance and energy forecasting.

Unit IV - Real-Time and Stream Analytics - (07 Hours)

Stream processing architectures, Apache Kafka and Spark Streaming, complex event processing (CEP), windowing operations, edge analytics concepts, latency and fault-tolerance in streaming systems, use cases in smart cities and industrial IoT

Unit V - Tools and Applications for IoT Analytics - (07 Hours)

IoT analytics platforms: AWS IoT, Azure IoT Hub, Google Cloud IoT, data storage solutions including InfluxDB, MongoDB, and time-series databases, dashboard tools such as Grafana and Kibana, integration with visualization libraries and BI tools, ethics, security, and privacy concerns in IoT analytics.

Learning Resources

Text Books

1. Bahga, A. & Madisetti, V., Internet of Things: A Hands-on Approach, Universities Press
2. Gupta, P. & Dey, N., Internet of Things and Data Analytics Handbook, CRC Press
3. Pang, Z. et al., Data-Driven Solutions for Smart Cities, Springer

Reference Books:

1. Apache Kafka: https://kafka.apache.org
2. Apache Spark Streaming: https://spark.apache.org/streaming
3. AWS IoT Analytics: https://aws.amazon.com/iot-analytics
4. Grafana Documentation: https://grafana.com/docs
5. Microsoft Azure IoT: https://azure.microsoft.com/en-in/services/iot-hub

SWAYAM / MOOC / YouTube Links

1. Introduction to Internet of Things – by Prof. Sudip Misra, IIT Kharagpur: https://nptel.ac.in/courses/10
2. Big Data Computing – by Prof. Rajiv Misra, IIT Patna: https://nptel.ac.in/courses/106/105/106105175
3. Cloud Computing – by Prof. Rajkumar Buyya, University of Melbourne: https://nptel.ac.in/courses/10

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Course)		
SEM-581-DS - Technical Seminar - I		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 04 Hours/Week	02	**Term Work** : 25 Marks **Oral/Presentation** : 25 Marks

Course Description:

The seminar aims to enhance students' research, presentation, and critical thinking skills, preparing them for advanced academic pursuits and professional careers. Technical Seminars will provide students with the opportunity and support to improve their self-study skills using modern information technologies and apply new knowledge and skills in practice, including in new areas.

Course Objectives: Upon successful completion of this course, students will be able to:

- **Deepen Technical Knowledge:** To enable students to explore a specialized topic within Computer Engineering beyond the regular curriculum, fostering in-depth understanding.
- **Develop Research Skills:** To provide practical experience in identifying, acquiring, evaluating, and synthesizing information from various technical sources (research papers, standards, technical reports).
- **Enhance Communication Skills:** To cultivate effective oral and visual presentation skills, enabling students to articulate complex technical concepts clearly and concisely to a knowledgeable audience.
- **Foster Critical Thinking:** To encourage students to critically analyze existing research, identify challenges, propose solutions, and engage in constructive discussions.
- **Promote Independent Learning:** To encourage self-directed learning and the ability to stay updated with emerging technologies and research trends.
- **Prepare for Thesis/Dissertation:** To serve as a foundational step for the Master's thesis/dissertation, allowing students to explore potential research areas.

Course Outcomes: After successful completion of the course, learner will be able to:

- **CO1 : Formulate** the goals and objectives of scientific research;
- **CO2 :** Search, evaluate and **analyze** information about the achievements of science and technology in the target area and beyond;
- **CO3 : Interpret** data from different fields of science and technology;
- **CO4 : Build** the logic of reasoning and statements;
- **CO5 : Create**, design and edit text documents in accordance with the requirements of the organization or publisher;

Guidelines for Seminar

- **Responsibility of the students:**
  - The Seminar should be carried out individually by each student.
  - A student should identify the area or topics in recent trends and developments in consultation with the guide.
  - A student should report to his/her respective guide regularly (at least once in a week) and report the progress of the seminar work.
  - A student should follow the timelines and deadlines and inform the supervisor in case of any difficulty/delay.
  - Students should maintain the record of all the meetings, remarks given by guide/reviewers and progress of the work in the diary. The diary must be presented during each review presentation to the reviewers.
  - A student should conduct the research ethically, adhere to the academic integrity standards, and cite sources whenever using any existing results
  - A student should incorporate constructive feedback to improve the quality and rigor of the research work towards seminar.
  - For final examination, students should complete the Seminar Report in all aspects including formatting and citation.
  - Each student should prepare the report, get it approved by his/her guide and submit the duly signed copy within the deadline.
  - A student should invest time and effort in preparing for seminar presentations and the oral presentation of the seminar.
- **Topic Selection**
  - Relevance: Topics must be directly related to Computer Engineering, encompassing current research trends, emerging technologies, advanced concepts, or interdisciplinary applications.
  - Scope: The topic should be sufficiently focused to allow for in-depth exploration within the seminar timeframe, yet broad enough to demonstrate a comprehensive understanding. Avoid overly narrow or excessively broad topics.
  - Novelty (Desired): While not strictly a research paper, students are encouraged to explore topics that have recent advancements, open problems, or areas where their unique insights can be presented. Avoid merely summarizing introductory textbook material.
  - Guide / Supervisor Approval: Each student must select a seminar topic in consultation with and obtain approval from an assigned faculty supervisor. The supervisor will guide the student in refining the topic and identifying relevant resources.
  - Examples of Broad Areas: Artificial Intelligence/Machine Learning, Data Science & Big Data, Cybersecurity, Cloud Computing, Internet of Things (IoT), Computer Networks, Software Engineering, High-Performance Computing, Embedded Systems, Computer Vision, Natural Language Processing, Blockchain, Quantum Computing.

- **Seminar Structure and Deliverable :** The technical seminar typically involves the following stages and deliverable
  - Topic Proposal (2-3 weeks after topic approval)
  - A concise document (1-2 pages) outlining:
    * Proposed Seminar Title
    * Brief Description/Abstract of the Topic
    * Motivation and Relevance to Computer Engineering
    * Preliminary List of Key References (at least 5-7 reputable sources)
    * Tentative Scope and Outline of the Presentation
    * Submission: To the faculty supervisor for approval.
    * Literature Review and Research (Ongoing): Sources: Students must primarily rely on peer-reviewed academic sources (IEEE Xplore, ACM Digital Library, SpringerLink, arXiv, Google Scholar), reputable conference proceedings, and established industry standards. Wikipedia and unverified blogs are generally not acceptable as primary sources.
    * Critical Analysis: Beyond mere summarization, students are expected to critically analyze the literature, identifying different approaches, their advantages/disadvantages, open issues, and potential future directions.
    * Note-Taking & Organization: Maintain systematic notes and organize research material effectively.
- **Seminar Report (Due 2-3 weeks before presentation):**
  - A written report (typically 20-25 pages, excluding references and appendices) detailing the seminar content.
  - Format: Follow a professional academic paper format (e.g., IEEE transaction style).
  - Sections:
    * Abstract: A concise summary of the seminar topic and key findings.
    * Introduction: Background, motivation, problem statement (if applicable), and outline of the report.
    * Literature Review/Background: Detailed discussion of relevant concepts, theories, and existing work.
    * Core Content: In-depth exploration of the chosen topic, presenting different methodologies, architectures, algorithms, or challenges as relevant.
    * Analysis/Discussion: Critical evaluation of the presented material, comparing different approaches, discussing implications, and identifying gaps.
    * Future Trends/Conclusion: Summarization of key takeaways, potential future directions, and concluding remarks.
    * References: A comprehensive list of all cited sources, properly formatted.

* Appendices (Optional): Supplementary material if necessary.

- **Oral Presentation :**
  - Duration: Typically 25-30 minutes for presentation, followed by 10-15 minutes for Q&A.
  - Audience: Faculty members, peers, and potentially other interested individuals.
  - Content: The presentation should effectively convey the key aspects of the seminar topic. It should not simply be a reading of the report.
  - Visual Aids: High-quality presentation slides (e.g., PowerPoint, Google Slides, LaTeX Beamer) are mandatory. Slides should be clear, concise, visually appealing, and support the oral delivery. Avoid excessive text on slides.
  - Delivery: Clear articulation, confident posture, good eye contact, and appropriate pace. Practice the presentation thoroughly.
  - Q&A Session: Be prepared to answer questions from the audience on all aspects of the seminar topic. Demonstrate a strong understanding and ability to defend your perspectives.
- **Evaluation Criteria :** The technical seminar will be evaluated based on the following criteria:
  - **Topic Selection and Scope (10%):** Relevance, timeliness, and appropriate depth of the chosen topic. Clarity and focus of the topic proposal.
  - **Literature Review and Research (25%):** Breadth and depth of literature surveyed. Quality and credibility of sources used. Critical analysis and synthesis of information.
  - **Seminar Report/Paper (30%):** Clarity, organization, and logical flow of content. Technical accuracy and depth of discussion. Adherence to academic writing standards (grammar, spelling, formatting, referencing). Originality in synthesis and critical insights. Absence of plagiarism.
  - **Oral Presentation (35%):** Content: Clarity, completeness, and accuracy of the presented material. Organization: Logical flow, effective use of time. Visual Aids: Quality, clarity, and effectiveness of slides. Delivery: Confidence, clarity of speech, enthusiasm, engagement with the audience. Q&A: Ability to answer questions accurately, comprehensively, and confidently.

Learning Resources

Text Books

1. "Engineering Communication" by Charles W. Knisely & Karin I. Knisely
2. "Technical Communication: Principles and Practice" by Meenakshi Raman & Sangeeta Sharma
3. "The Craft of Scientific Presentations" by Michael Alley

NPTEL Courses

1. https://nptel.ac.in/courses/109/106/109106180/
2. https://www.udemy.com/course/technical-writing/
3. https://www.edx.org/course/writing-in-the-sciences

Savitribai Phule Pune University, Pune

Maharashtra, India

M. E. - Data Science (DS) - 2025 Pattern

Semester III

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Pattern)		
RM-631-DS - Research Methodology		
Teaching Scheme	**Credits**	**Examination Scheme**
Theory : 04 Hours/Week	04	**CCE** : 50 Marks **End-Semester**: 50 Marks

Prerequisite::

1. Sound fundamentals of the core engineering/science domain (since research is often domain-specific).

Course Objectives: The course aims to:

1. To understand the philosophy of research in general
2. To learn the methodology to conduct the Literature Survey
3. To acquaint with the tools, techniques, and processes of doing research
4. To learn the effective report writing skills and allied documentations
5. To become aware of the ethics in research, academic integrity and plagiarism

Course Outcomes: Upon successful completion of this course, students will be able to:
CO1: Formulate research problems with clear objectives, hypotheses, and scope.
CO2: Design appropriate research methodologies using qualitative, quantitative, or mixed approaches.
CO3: Conduct literature reviews, identify research gaps, and critically evaluate scholarly articles.
CO4: Apply statistical and analytical tools for data collection, analysis, and interpretation.
CO5: Prepare technical reports, research papers, and project proposals adhering to ethical and academic standards.

Course Contents

Unit I - Foundations of Research - (08 Hours)

Meaning, Objectives, Motivation. Concept of theory, deductive and inductive theory. Characteristics of scientific method – Understanding the language of research – Concept, Construct, Definition, Variable. Research Process. Problem Identification & Formulation – Research Question – Investigation Question – Measurement Issues – Hypothesis – Qualities of a Good Hypothesis –Null Hypothesis &; Alternative Hypothesis. Hypothesis Testing – Logic &; Importance

Unit II - Research Design - (08 Hours)

Concept and Importance in Research – Features of a good research design – Exploratory Research Design – concept, types, and uses, Descriptive Research Designs – concept, types, and uses. Experimental Design: Concept of Independent & Dependent variables. Qualitative and Quantitative Research: Qualitative research – Quantitative research – Concept of measurement, causality, generalization, and replication. Merging the two approaches

Unit III - Sampling, Measurement and Data Analysis - (08 Hours)

Concepts of Statistical Population, Sample, Sampling Frame, Sampling Error, Sample Size, Non-Response. Characteristics of a good sample. Probability Sample – Simple Random Sample, Systematic, Sample, Stratified Random Sample & Multi-stage Sampling. Determining the size of the sample – Practical considerations in sampling and sample size. Concept of measurement– what is measured?, Problems in measurement in research – Validity and Reliability. Levels of measurement – Nominal, Ordinal, Interval, Ratio. Data Analysis: Data Preparation – Univariate analysis (frequency tables, bar charts, pie charts, percentages). Basic principle of Analysis of Variance, ANOVA Technique, Setting up Analysis of Variance Table, short-cut method for oneway ANOVA, Coding method, Two-way ANOVA, ANOVA in Latin-square design, analysis of co-variance (ANCOVA), assumptions in ANCOVA. Academic Ethics: Plagiarism, exposure on anti-plagiarism tools.

Unit IV - Thesis writings, Paper Writing and IPR- (08 Hours)

Draft thesis, Layout of a Research Paper, paper communication, and publications, Journals, Impact factor of Journals, When and where to publish? Ethical issues related to publishing, Plagiarism, and Self-Plagiarism. Major contribution, outcome of the research, patent possibilities.

Unit V - Databases and Use of Tools - (08 Hours)

Use of Encyclopedias, Research Guides, Handbook etc., Academic Databases for Computer Science Discipline. Use of tools / techniques for Research: methods to search required information effectively, Reference Management Software like Zotero/Mendeley, Software for paper formatting like LaTeX/MS Office, Software for detection of Plagiarism.

Suggested Assignment

1.**Literature Review and Research Gap Identification:**

Select a research topic in your field of interest. Conduct a systematic literature review by collecting at least 15 relevant research papers. A report summarizing the key findings, methodologies, and conclusions from the papers, and clearly identify one or more gaps or open questions in the existing research.

2. **Research Proposal Writing:**

Write a complete research proposal including introduction, problem statement, literature review, research questions, methodology, and expected outcomes.

A formal research proposal document ready for submission or peer review.

3.**Presentation and Visualization of Research Findings:**

Create visualizations (charts, graphs) to represent data findings and prepare a research presentation summarizing the study.

A slide deck with clear visuals and a summary report explaining the findings.

Learning Resources

Text Books

1. C. R. Kothari and G. Garg, Research Methodology: Methods and Techniques, 4th ed. New Delhi, India: New Age International Publishers, 2019.

2. J. W. Creswell and J. D. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th ed. Thousand Oaks, CA, USA: SAGE Publications, 2017.

3. R. Kumar, Research Methodology: A Step-by-Step Guide for Beginners, 5th ed. Thousand Oaks, CA, USA: SAGE Publications, 2019.

Reference Books:

1. W. J. Goode and P. K. Hatt, Methods in Social Research. New York, NY, USA: McGraw-Hill, 1952.

2. Thomas, C. G. (2021). Research methodology and scientific writing (pp. 135-151). Thrissur: Springer.

SWAYAM / MOOC / YouTube Links

1. https://www.youtube.com/playlist?list=PLm-zueI9b64QGMcfn5Ckv_8W5Z1d3vMBY

2. https://onlinecourses.swayam2.ac.in/cec20_hs17/preview

3. https://onlinecourses.nptel.ac.in/noc23_ge36/preview

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Course)		
OJT-602-DS - Internship/On Job Training (IN/OJT)		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 10 Hours/Week	05	Term Work : 100 Marks

Course objectives :

1. To put theory into practice.and expand thinking and broaden the knowledge and skills acquired through course work in the field.
2. To relate to, interact with, and learn from current professionals in the field.
3. To understand and adhere to professional standards in the field.
4. To gain insight to professional communication including meetings, memos, reading, writing, public speaking, research, client interaction, input of ideas, and confidentiality.
5. To develop the initiative and motivation to be a self-starter and work independently.

Course Outcomes: Upon successful completion of this course, students will be able to:

- **Gain** practical experience within industry in which the internship is done.
- **Acquire** knowledge of the industry in which the internship is done.
- **Apply** knowledge and skills learned to classroom work.
- **Develop** and refine oral and written communication skills.
- **Acquire** the knowledge of administration, marketing, finance and economics.

Course Description:

1. Internship/On Job Training provide students the opportunity of hands-on experience that includes personal training, time and stress management, interactive skills, presentations, budgeting, marketing, liability and risk management, paperwork, equipment ordering, maintenance, responding to emergencies etc.
2. An internship is the phase of time for students when they are trained for their skills, they are good at, and it gives them a chance to apply their knowledge practically in industries
3. The internship can be carried out in any industry/R&D Organization/Research Institute/Institute of national repute/R&D Centre of Parent Institute.
4. The Department/college shall nominate a faculty to facilitate, guide and supervise students under internship.

Guidelines

- **Purpose:** Internships are designed to bridge the gap between academic learning and industry practice. They aim to provide hands-on experience, expose students to the industrial environment, develop technical and soft skills (communication, teamwork, problem-solving), and help in career exploration.
- **Internship Duration and Academic Credentials**
  - Student can take internship work in the form of Online/Offline mode from any of the Industry / Government Organization Internship Programmes approved by SPPU/AICTE/UGC portals
  - A intern is expected to spend 10 - 12 hours per week on Internship, Training will result in about 160-170 hours of total internship duration.
  - The minimum requirement regarding Internship duration should not be below 8 weeks
- **Type of Internship**
  - o Industry/Government Organization Internship: Working directly with a company or government body.
  - o Research Internship: Focused on research projects, often in collaboration with academic institutions or R&D labs.
  - o Innovation/Entrepreneurship: Working on developing new products, processes, or even starting a venture.
  - o Social Internship: Engaging in community-based projects.
- **Assessment Details (TW and Practical)**
  - Term work for 100 marks
  - A daily log submitted by the student and a work log signed by the office HoDs where the student has interned will be considered towards the TW marking.
- **Indicative list of areas for OJT**
  - Trade and Agriculture
  - Economy & Banking Financial Services and Insurance
  - Logistics, Automotive & Capital Goods
  - Fast Moving Consumer Goods & Retail
  - Information Technology/Information Technology Enabled Services & Electronics
  - Handcraft, Art, Design & Music
  - Healthcare & Life Science
  - Sports, Wellness and Physical Education
  - Tourism & Hospitality
  - Digitization & Emerging Technologies (Internet of Things / Artificial Intelligence / Machine

  - Learning / Deep Learning / Augmented Reality / Virtual Reality etc.)
  - Humanitarian, Public Policy and Legal Services
  - Communication
  - Education
  - Sustainable Development
  - Environment
  - Commerce, Medium and Small-Scale Industries

- **Faculty Supervision:** Students are usually assigned an internal faculty guide/mentor who supervises their internship activities. This faculty member acts as a teacher, mentor, and critic, and ensures the internship aligns with academic goals. External Supervision: In many cases, an external expert from the host organization also guides the student.

- **Documentation and Reporting:**

  - Joining Report: To be submitted within a specified time frame (e.g., one week from joining).
  - Daily/Periodical Diary: Students are often required to maintain a daily or weekly record of their observations, work, and learning.
  - Internship Report: A comprehensive report detailing the work done, learning outcomes, and achievements during the internship. This report needs to be duly signed by the company official and faculty mentor.
  - Completion Certificate: Issued by the host organization upon successful completion.

- **Evaluation :**

  - Evaluation is typically done by the institute, often within a short period after the internship ends.
  - It may involve presentations, viva-voce examinations, and assessment of the internship report and daily diary.
  - Performance-based feedback from the industry mentor is usually a key component.

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Course)		
SEM-603-DS - Technical Seminar - II		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 08 Hours/Week	04	**Term Work:**50 Marks **Oral/Presentation:** 50 Marks

Course Description:

- Research Project seminar is the first stage of work on a master's thesis. During this course, students gain experience in the field of intellectual property and research ethics. They conduct patent searches and analyze related works to study the current state of the target area.
- Work on the "Research Project seminar" is carried out on the basis of the research and training laboratories of the Institute and the Scientific Library of the Institute/University and in close cooperation with the student's supervisor.
- The aim of the "Research Project Seminar " is to prepare for the implementation of the Final Project and for master's thesis defense. It includes finding or developing methods and tools to solve a stated problem, taking into account the latest research and trends; clarification of requirements for the object under development; planning experiments and tests to prove the effectiveness of the proposed solution

Course Objectives: Upon successful completion of this course, students will be able to:

- To provide students with the opportunity and support to improve their self-study skills using modern information technologies and apply new knowledge and skills in practice, including in new areas.
- To raise student's awareness in advanced methods of research and mastering the skills to apply them.
- Teach students to find and critically analyze sources of information.
- Develop their ability to build logic of reasoning and statements based on the interpretation of data combined from various fields of science and technology, to make judgments based on incomplete data.
- Improve the student's academic writing experience.

Course Outcomes: After successful completion of the course, learner will be able to:

- **CO1 -** Gain fundamental concepts and categories in the field of scientific research- ways of organizing and planning research
- **CO2** - Advanced information technologies allowing us to **acquire** new knowledge in various fields
- **CO3 - Learn** features of the technical and scientific style of writing texts

- **CO4 -** Evaluation criteria and methods of handling incomplete data

By the end of the course, students will be able to:

- formulate the goals and objectives of scientific research;
- search, evaluate and analyze information about the achievements of science and technology in the target area and beyond;
- interpret data from different fields of science and technology;
- to build the logic of reasoning and statements;
- write a text in a scientific or scientific and technical style, use the appropriate vocabulary;
- create, design and edit text documents in accordance with the requirements of the organization or publisher;
- plan a pilot study
  - methods of planning scientific research, taking into account the peculiarities of the professional area.
  - methods of collecting and analyzing information on the achievements of science and technology in the target area and beyond.
  - proficiency in preparing publications on the topic of research
  - experience in data integration from different fields of science and technology and building evidence-based judgments.
  - methods of planning an experiment, taking into account the peculiarities of the field of professional activity.

Responsibility of the students:

- This Seminar should be carried out individually by each student based on the research project.
- A student should identify the area or topics in recent trends and developments in consultation with the guide
- A student should report to his/her respective guide regularly (at least once in a week) and report the progress of the seminar work.
- A student should follow the timelines and deadlines and inform the supervisor in case of any difficulty/delay.
- Students should maintain the record of all the meetings, remarks given by guide/reviewers and progress of the work in the project diary. The project diary must be presented during each review presentation to the reviewers.

- A student should conduct the research ethically, adhere to the academic integrity standards, and cite sources whenever using any existing results
- A student should Incorporate constructive feedback to improve the quality and rigor of the research
- For final examination, students should complete the Seminar Report in all aspects including formatting and citation.
- Each student should prepare the report, get it approved by his/her guide and submit the duly signed copy within the deadline.
- A student should invest time and effort in preparing for seminar presentations and the oral presentation of the seminar
- **Topic Selection**
  - Relevance: Topics must be directly related to Computer Engineering, encompassing current research trends, emerging technologies, advanced concepts, or interdisciplinary applications.
  - Scope: The topic should be sufficiently focused to allow for in-depth exploration within the seminar timeframe, yet broad enough to demonstrate a comprehensive understanding. Avoid overly narrow or excessively broad topics.
  - Novelty (Desired): While not strictly a research paper, students are encouraged to explore topics that have recent advancements, open problems, or areas where their unique insights can be presented. Avoid merely summarizing introductory textbook material.
  - Guide / Supervisor Approval: Each student must select a seminar topic in consultation with and obtain approval from an assigned faculty supervisor. The supervisor will guide the student in refining the topic and identifying relevant resources.
  - Examples of Broad Areas: Artificial Intelligence/Machine Learning, Data Science & Big Data, Cybersecurity, Cloud Computing, Internet of Things (IoT), Computer Networks, Software Engineering, High-Performance Computing, Embedded Systems, Computer Vision, Natural Language Processing, Blockchain, Quantum Computing.
- **Seminar Structure and Deliverable :** The technical seminar typically involves the following stages and deliverable
  - Topic Proposal (2-3 weeks after topic approval)
  - A concise document (1-2 pages) outlining:
    * Proposed Seminar Title
    * Brief Description/Abstract of the Topic
    * Motivation and Relevance to Computer Engineering
    * Preliminary List of Key References (at least 5-7 reputable sources)
    * Tentative Scope and Outline of the Presentation

    - Submission: To the faculty supervisor for approval.
    - Literature Review and Research (Ongoing): Sources: Students must primarily rely on peer-reviewed academic sources (IEEE Xplore, ACM Digital Library, SpringerLink, arXiv, Google Scholar), reputable conference proceedings, and established industry standards. Wikipedia and unverified blogs are generally not acceptable as primary sources.
    - Critical Analysis: Beyond mere summarization, students are expected to critically analyze the literature, identifying different approaches, their advantages/disadvantages, open issues, and potential future directions.
    - Note-Taking & Organization: Maintain systematic notes and organize research material effectively.

- **Seminar Report (Due 2-3 weeks before presentation):**
  - A written report (typically 20-25 pages, excluding references and appendices) detailing the seminar content.
  - Format: Follow a professional academic paper format (e.g., IEEE transaction style).
  - Sections:
    - Abstract: A concise summary of the seminar topic and key findings.
    - Introduction: Background, motivation, problem statement (if applicable), and outline of the report.
    - Literature Review/Background: Detailed discussion of relevant concepts, theories, and existing work.
    - Core Content: In-depth exploration of the chosen topic, presenting different methodologies, architectures, algorithms, or challenges as relevant.
    - Analysis/Discussion: Critical evaluation of the presented material, comparing different approaches, discussing implications, and identifying gaps.
    - Future Trends/Conclusion: Summarization of key takeaways, potential future directions, and concluding remarks.
    - References: A comprehensive list of all cited sources, properly formatted.
    - Appendices (Optional): Supplementary material if necessary.

- **Oral Presentation :**
  - Duration: Typically 25-30 minutes for presentation, followed by 10-15 minutes for Q&A.
  - Audience: Faculty members, peers, and potentially other interested individuals.
  - Content: The presentation should effectively convey the key aspects of the seminar topic. It should not simply be a reading of the report.
  - Visual Aids: High-quality presentation slides (e.g., PowerPoint, Google Slides, LaTeX Beamer) are mandatory. Slides should be clear, concise, visually appealing, and support the oral delivery. Avoid excessive text on slides.

- Delivery: Clear articulation, confident posture, good eye contact, and appropriate pace. Practice the presentation thoroughly.
- Q&A Session: Be prepared to answer questions from the audience on all aspects of the seminar topic. Demonstrate a strong understanding and ability to defend your perspectives.

- **Evaluation Criteria :** The technical seminar will be evaluated based on the following criteria:
  - **Topic Selection and Scope (10%):** Relevance, timeliness, and appropriate depth of the chosen topic. Clarity and focus of the topic proposal.
  - **Literature Review and Research (25%):** Breadth and depth of literature surveyed. Quality and credibility of sources used. Critical analysis and synthesis of information.
  - **Seminar Report/Paper (30%):** Clarity, organization, and logical flow of content. Technical accuracy and depth of discussion. Adherence to academic writing standards (grammar, spelling, formatting, referencing). Originality in synthesis and critical insights. Absence of plagiarism.
  - **Oral Presentation (35%):** Content: Clarity, completeness, and accuracy of the presented material. Organization: Logical flow, effective use of time. Visual Aids: Quality, clarity, and effectiveness of slides. Delivery: Confidence, clarity of speech, enthusiasm, engagement with the audience. Q&A: Ability to answer questions accurately, comprehensively, and confidently.

Learning Resources

Text Books

1. Kennett, B. (2014). Planning and managing scientific research. ANU Press. https://www.jstor.org/stabl (free access)
2. Sirotinina, N. (2012). History and methodology of computer science. Siberian Federal University. Tomsk: TPU Publishing House.
3. Moore, N. (2006). How to do research: a practical guide to designing and managing research projects. Facet publishing.

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Course)		
RPR-604-DS - Research Project Stage - I		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 18 Hours/Week	09	**Term Work :** 25 Marks **Oral/ Presentation :** 25 Marks

Course Description:

The master's degree culminates in a research project of the student's own design. This research project is documented by a final research report or dissertation. The student's work is guided by an academic supervisor. Students are expected to choose real-world contemporary problem and apply the engineering principles learned, to solve the problem through building prototypes or simulations or writing codes or establishing processes/synthesis/correlations etc.

Students are expected to construct a research project that includes original research, deliberate and well considered methodological choices, and shows relevance to significant conversations within the discipline. The dissertation should represent the very best research and analysis a student can produce.

Course Objectives: Upon successful completion of this course, students will be able to:

1. Demonstrate an ability to plan a research project, such as is required in a research proposal prior to the launch of their work
2. Demonstrate an ability to comply with ethical, safety, and documentation processes appropriate to their project
3. Demonstrate expert knowledge in the subject of their research project, such as through a integrated literature survey
4. Demonstrate expert knowledge in the research methods appropriate to generating reliable data for their research questions
5. Demonstrate the ability to manage projects and to make constructive use of expertise associated with their project, while working as an independent learner
6. Demonstrate an ability to relate their original data to existing literature, or to create an novel synthesis of existing materials

Course Outcomes:

After successful completion of the course, learner will be able to:

- CO 1 : Demonstrate how to search the existing literature to gather information about a specific problem or domain.
- CO 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.

- CO 3 : Evaluate various solution techniques to determine the most feasible solution within given constraints for the chosen dissertation problem.
- CO 4 : Apply software engineering principles related to requirements gathering and design to produce relevant documentation.
- CO 5 : Write a dissertation report that details the research problem, objectives, literature review, and solution architecture.

Guidelines for Research Project

1. **General Guidelines :**

   (a) 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.

   (b) Student is expected to complete the following activities in Phase-I:

      i. Literature survey
      ii. Problem Definition
      iii. Motivation for study and Objectives
      iv. Preliminary design / feasibility / modular approaches
      v. Design of the research project

Phase 1: Informal conversations

Students are strongly encouraged to discuss possible research project ideas with the internal guide , fellow students, and other research professionals. All research projects begin with open-ended conversations and scoping exercises.

Phase 2: Identify topic

The first formal step in the module involves identifying a preliminary project title and writing an abstract of no more than 500 words. Writing an abstract for a research proposal or for completed research work is an important transferable skill.

The project title is understood to be provisional. Supervisors/guide will be assigned to students after the project title/ abstract forms have been submitted. The main responsibilities of the supervisor/guide are to assist the student with project management and to advise the student on criteria for assessment. It is a good idea to discuss a time line for your project with your supervisor/guide, and to establish a definite timetable.

Phase 3: Project proposal

The proposal should reflect a student's best effort. At the same time, we recognize research often raises new questions. Some redefinitions of topics and titles is common later in the research process.

Students should keep their supervisors up to date on these developments, and they can expect a reasonable amount of adaptation.

Phase 4: Term-1 research

Students are expected to commit substantial time during the term to their research project. The principal form of academic input for the research project normally comes through discussions with the designated supervisor. The majority of these meetings should be face-to-face, either in person or via video- or audio-conferencing technology. Students are expected to respect these periods of absence and plan their needs accordingly.

Phase 5: Submit project report

The project report with the specific due date must be submitted to department.

Additional Information

- **Research notebook :** Students are strongly advised to maintain a research notebook, either digital or paper, and to keep this up to date. A research notebook can prove useful should examiners query research methods, research integrity, or research process.
- **Preventing data loss:** Protect yourself against loss of research material and writing by maintaining a system for secure, redundant, up-to-date back-up of research material and writing. Loss cannot be accepted as a reason for failing to meet a deadline. A copy of written notebooks can be stored by supervisors for the duration of the project. Loss of project materials through accidents and theft have occurred in the past; these have had devastating effects on the unprepared. All students are warned to create redundancies to protect their project from similar calamities.
- **Citation format :** The style must be clear, explicit, and meaningful. As a recommendation, students should use a style frequently used in the literature relevant to their research project. Most journals have style guides in their notes to contributors. Students should discuss options with their supervisors, and they should keep in mind that efficient citation is one element in the criteria for assessment.

Savitribai Phule Pune University, Pune

Maharashtra, India

M. E. - Data Science (DS) - 2025 Pattern

Semester IV

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Course)		
SEM-651-DS - Technical Seminar - III		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 08 Hours/Week	04	**Term Work:**50 Marks **Oral/Presentation:** 50 Marks

Course Description:

The seminar aims to enhance students' research, presentation, and critical thinking skills, preparing them for advanced academic pursuits and professional careers. Technical Seminars will provide students with the opportunity and support to improve their self-study skills using modern information technologies and apply new knowledge and skills in practice, including in new areas.

Course Objectives: Upon successful completion of this course, students will be able to:

- **Deepen Technical Knowledge:** To enable students to explore a specialized topic within Computer Engineering beyond the regular curriculum, fostering in-depth understanding.
- **Develop Research Skills:** To provide practical experience in identifying, acquiring, evaluating, and synthesizing information from various technical sources (research papers, standards, technical reports).
- **Enhance Communication Skills:** To cultivate effective oral and visual presentation skills, enabling students to articulate complex technical concepts clearly and concisely to a knowledgeable audience.
- **Foster Critical Thinking:** To encourage students to critically analyze existing research, identify challenges, propose solutions, and engage in constructive discussions.
- **Promote Independent Learning:** To encourage self-directed learning and the ability to stay updated with emerging technologies and research trends.
- **Prepare for Thesis/Dissertation:** To serve as a foundational step for the Master's thesis/dissertation, allowing students to explore potential research areas.

Course Outcomes: After successful completion of the course, learner will be able to:

- **CO1 : Formulate** the goals and objectives of scientific research;
- **CO2 :** Search, evaluate and **analyze** information about the achievements of science and technology in the target area and beyond;
- **CO3 : Interpret** data from different fields of science and technology;
- **CO4 : Build** the logic of reasoning and statements;
- **CO5 : Create**, design and edit text documents in accordance with the requirements of the organization or publisher;

Guidelines for Seminar

- **Responsibility of the students:**
  - The Seminar should be carried out individually by each student.
  - A student should identify the area or topics in recent trends and developments in consultation with the guide.
  - A student should report to his/her respective guide regularly (at least once in a week) and report the progress of the seminar work.
  - A student should follow the timelines and deadlines and inform the supervisor in case of any difficulty/delay.
  - Students should maintain the record of all the meetings, remarks given by guide/reviewers and progress of the work in the diary. The diary must be presented during each review presentation to the reviewers.
  - A student should conduct the research ethically, adhere to the academic integrity standards, and cite sources whenever using any existing results
  - A student should incorporate constructive feedback to improve the quality and rigor of the research work towards seminar.
  - For final examination, students should complete the Seminar Report in all aspects including formatting and citation.
  - Each student should prepare the report, get it approved by his/her guide and submit the duly signed copy within the deadline.
  - A student should invest time and effort in preparing for seminar presentations and the oral presentation of the seminar.
- **Topic Selection**
  - Relevance: Topics must be directly related to Computer Engineering, encompassing current research trends, emerging technologies, advanced concepts, or interdisciplinary applications.
  - Scope: The topic should be sufficiently focused to allow for in-depth exploration within the seminar timeframe, yet broad enough to demonstrate a comprehensive understanding. Avoid overly narrow or excessively broad topics.
  - Novelty (Desired): While not strictly a research paper, students are encouraged to explore topics that have recent advancements, open problems, or areas where their unique insights can be presented. Avoid merely summarizing introductory textbook material.
  - Guide / Supervisor Approval: Each student must select a seminar topic in consultation with and obtain approval from an assigned faculty supervisor. The supervisor will guide the student in refining the topic and identifying relevant resources.
  - Examples of Broad Areas: Artificial Intelligence/Machine Learning, Data Science & Big Data, Cybersecurity, Cloud Computing, Internet of Things (IoT), Computer Networks, Software Engineering, High-Performance Computing, Embedded Systems, Computer Vision, Natural Language Processing, Blockchain, Quantum Computing.

- **Seminar Structure and Deliverable :** The technical seminar typically involves the following stages and deliverable
  - Topic Proposal (2-3 weeks after topic approval)
  - A concise document (1-2 pages) outlining:
    * Proposed Seminar Title
    * Brief Description/Abstract of the Topic
    * Motivation and Relevance to Computer Engineering
    * Preliminary List of Key References (at least 5-7 reputable sources)
    * Tentative Scope and Outline of the Presentation
    * Submission: To the faculty supervisor for approval.
    * Literature Review and Research (Ongoing): Sources: Students must primarily rely on peer-reviewed academic sources (IEEE Xplore, ACM Digital Library, SpringerLink, arXiv, Google Scholar), reputable conference proceedings, and established industry standards. Wikipedia and unverified blogs are generally not acceptable as primary sources.
    * Critical Analysis: Beyond mere summarization, students are expected to critically analyze the literature, identifying different approaches, their advantages/disadvantages, open issues, and potential future directions.
    * Note-Taking & Organization: Maintain systematic notes and organize research material effectively.
- **Seminar Report (Due 2-3 weeks before presentation):**
  - A written report (typically 20-25 pages, excluding references and appendices) detailing the seminar content.
  - Format: Follow a professional academic paper format (e.g., IEEE transaction style).
  - Sections:
    * Abstract: A concise summary of the seminar topic and key findings.
    * Introduction: Background, motivation, problem statement (if applicable), and outline of the report.
    * Literature Review/Background: Detailed discussion of relevant concepts, theories, and existing work.
    * Core Content: In-depth exploration of the chosen topic, presenting different methodologies, architectures, algorithms, or challenges as relevant.
    * Analysis/Discussion: Critical evaluation of the presented material, comparing different approaches, discussing implications, and identifying gaps.
    * Future Trends/Conclusion: Summarization of key takeaways, potential future directions, and concluding remarks.
    * References: A comprehensive list of all cited sources, properly formatted.

* Appendices (Optional): Supplementary material if necessary.

- **Oral Presentation :**
  - Duration: Typically 25-30 minutes for presentation, followed by 10-15 minutes for Q&A.
  - Audience: Faculty members, peers, and potentially other interested individuals.
  - Content: The presentation should effectively convey the key aspects of the seminar topic. It should not simply be a reading of the report.
  - Visual Aids: High-quality presentation slides (e.g., PowerPoint, Google Slides, LaTeX Beamer) are mandatory. Slides should be clear, concise, visually appealing, and support the oral delivery. Avoid excessive text on slides.
  - Delivery: Clear articulation, confident posture, good eye contact, and appropriate pace. Practice the presentation thoroughly.
  - Q&A Session: Be prepared to answer questions from the audience on all aspects of the seminar topic. Demonstrate a strong understanding and ability to defend your perspectives.
- **Evaluation Criteria :** The technical seminar will be evaluated based on the following criteria:
  - **Topic Selection and Scope (10%):** Relevance, timeliness, and appropriate depth of the chosen topic. Clarity and focus of the topic proposal.
  - **Literature Review and Research (25%):** Breadth and depth of literature surveyed. Quality and credibility of sources used. Critical analysis and synthesis of information.
  - **Seminar Report/Paper (30%):** Clarity, organization, and logical flow of content. Technical accuracy and depth of discussion. Adherence to academic writing standards (grammar, spelling, formatting, referencing). Originality in synthesis and critical insights. Absence of plagiarism.
  - **Oral Presentation (35%):** Content: Clarity, completeness, and accuracy of the presented material. Organization: Logical flow, effective use of time. Visual Aids: Quality, clarity, and effectiveness of slides. Delivery: Confidence, clarity of speech, enthusiasm, engagement with the audience. Q&A: Ability to answer questions accurately, comprehensively, and confidently.

Learning Resources

Text Books

1. "Engineering Communication" by Charles W. Knisely & Karin I. Knisely
2. "Technical Communication: Principles and Practice" by Meenakshi Raman & Sangeeta Sharma
3. "The Craft of Scientific Presentations" by Michael Alley

NPTEL Courses

1. https://nptel.ac.in/courses/109/106/109106180/
2. https://www.udemy.com/course/technical-writing/
3. https://www.edx.org/course/writing-in-the-sciences

Savitribai Phule Pune University Master of Engineering - Data Science (2025 Course)		
RPR-652-DS - Research Project Stage-II		
Teaching Scheme	**Credits**	**Examination Scheme**
Practical: 36 Hours/Week	18	**Term Work:**150 Marks **Oral/ Presentation :** 50 Marks

Prerequisite : Research Project Stage-I

Course Objectives: Upon successful completion of this course, students will be able to:

- **Demonstrate** an ability to plan a research project, such as is required in a research proposal prior to the launch of their work
- **Ability** to manage projects and to make constructive use of expertise associated with their project, while working as an independent learner
- **Ability** to relate their original data to existing literature, or to create an novel synthesis of existing materials
- **Identify** and **formulate** a problem of research interest in the chosen area of computing.

Course Outcomes: After successful completion of the course, learner will be able to:

1. **CO1 : Undertake** independent research that makes an original contribution to knowledge, or produces a novel synthesis of existing materials relevant to significant conversations in the discipline
2. **CO2 : Plan** their project in advance, using a proposal to describe their undertaking, describe how it will be managed, and reflect upon its value
3. **CO3 : Relate** their original research to existing literature on the subject and relate their work to general themes in their relevant scholarly literature
4. **CO4 : Assemble** their rationale, methods, findings, and analysis into a substantial piece of writing that presents a clear thesis and a cohesive evidence-based argument or analysis
5. **CO5 : Reflect** on the strengths and weaknesses of their research and methodology, understanding how they might improve their efforts in future work

Guidelines for Research Project

- **General Guidelines**
  - The student shall consolidate and complete the remaining part of the research work started in Semester III. This will consist of Selection of Technology, Installations, implementations, testing, Results, measuring performance, discussions using data tables per parameter considered for the improvement with existing/known algorithms/systems, comparative analysis, validation of results and conclusions.

- The student shall prepare the duly certified final report of dissertation in standard format for satisfactory completion of the work by the concerned guide and head of the Department/Institute.
- The students are expected to validate their study undertaken by publishing it at standard platforms.
- The investigations and findings need to be validated appropriately at standard platforms like conference and/or peer reviewed journal.
- The student has to exhibit continuous progress through regular reporting and presentations and proper documentation of the frequency of the activities in the sole discretion of the PG coordination/Head of the department. The continuous assessment of the progress needs to be documented unambiguously.
- Supervisor Interaction: Minimum one meeting per week.
- Logbook: Maintain a record of work progress and supervisor comments.
- Ethics: No plagiarism, false results, or unethical practices allowed.
- Backup: Keep source code, datasets, and reports backed up securely.
- Submission Format: Soft copy (PDF) + Hard copy as per institute norms.

- **Key Components:**
  - **Implementation**
    - Complete development/simulation/testing of the system or model.
    - Ensure correctness, efficiency, and validation of results.
  - **Results & Analysis**
    - Include experimental setup, datasets used, performance metrics.
    - Graphs, tables, and comparison with existing techniques.
    - Highlight key findings and their significance.
  - **Conclusion and Future Work**
    - Summarize outcomes, contributions, and applications.
    - Suggest extensions or improvements for future research.
  - **Paper Publication**
    - At least one paper (optional/encouraged) in peer-reviewed conference/journal.
    - Attach publication/proof as appendix (if available).
  - **Final Report Format**
    - Revised version of Stage 1 report with added implementation, results, and conclusion chapters.
    - Maintain academic writing standards and include all necessary references.
  - **Plagiarism Report**

    - Final version must again be checked and should not exceed 15% similarity.

- **Evaluation Parameters**
    - Completeness and quality of implementation
    - Anaysis and originality of results
    - Qulity of documentation and adherence to format
    - Viv-voce performance and clarity of understanding
    - Contribution to knowledge or innovation

Task Force for Curriculum Design and Development

Programme Coordinator

Dr. Vandana Dhingra - Member, Board of Studies - Computer Engineering

Team Members for Course Design

Dr. Vandana Dhingra	Department of Technology, Savitribai Phule Pune University
Dr. Sunita Jahirabadkar	MKSSS's Cummins College of Engineering, Pune
Dr. Mininath R. Bendre	Pravara Rural Engineering College, Loni
Dr. Radhika Bhagwat	MKSSS's Cummins College of Engineering, Pune
Prof . Pranajali Deshpande	MKSSS's Cummins College of Engineering, Pune

Members of Board of Studies

Dr. Nilesh Uke	Chairman Board of Studies, Indira College of Engineering and Management Pune
Dr. Pramod Patil	Dean (S&T) , Dr. D.Y. Patil Institute of Technology, Pune
Dr. Amol Potgantwar	Sandip Institute of Technology and Research Center, Nasik
Dr. Sachin Babar	Sinhhgad Institute of Technology Lonavala
Dr. Dhananjay Kshirsagar	Sanjivani College of Engineering, Kopargaon
Dr. Balwant Sonkamble	Pune Institute of Computer Technology, Pune
Dr. Suhasini Itkar	PES's Modern College of Engineering, Pune
Dr. Dipti Patil	Cummins College of Engineering for Women, Pune
Dr. Sachin Sakhare	Vishwakarma Institute of Information Technology, Pune
Dr. Dipak Patil	R.H. Sapat college of Engineering, Nasik
Dr. Vandana Dhingra	Department of Technology, Savitribai Phule Pune University
Dr. Deepali Ujalambkar	AISSMS College of Engineering, Pune
Dr. Vaishali Vikhe	Pravara Rural Engineering College, Loni
Dr. Pradip Jawandhiya	PR Pote College of Engineering and Management, Amravati
Dr. Sandeep Deshmukh	TeemGenie, Pune

Chairman

Dr. Nilesh Uke - Board of Studies Computer Engineering

Savitribai Phule Pune University, Pune

Dean

Dr. Pramod Patil - Dean – Science and Technology

Savitribai Phule Pune University, Pune