Aliah University

Department of Mathematics and Statistics

Syllabus for

2-Year Masters in Data Science and Applied Statistics

(w.e.f. Academic Session 2025–26)

Program Objective

This syllabus aims to teach the basic concepts and theories of data science and applied statistics. It focuses on how to apply statistical theory in real-world situations. The program helps students learn statistical modeling, computing techniques, and machine learning to solve data-related problems. It combines theory with hands-on practice in programming, data visualization, and statistical analysis. Students will develop skills to understand and interpret data effectively. The syllabus also promotes critical thinking, problem-solving, and research skills. Graduates will be prepared for further studies and careers in areas like business analytics, healthcare, finance, artificial intelligence, and big data.

The Master of Data Science and Applied Statistics program is designed to equip students with a strong foundation in statistical theory, data analysis, and computational techniques essential for tackling real-world data challenges. The program aims to:

Program Outcomes (POs)

Code	**Description**
PO1	**Statistical and Mathematical Expertise:** Demonstrate advanced understanding of probability, mathematical foundations, and statistical theories to model real-world data problems.
PO2	**Data Management and Engineering:** Apply modern database principles, SQL, and data structures for efficient retrieval, cleansing, and transformation of large datasets.
PO3	**Computational Proficiency:** Utilize programming languages (Python, R) and relevant software libraries to implement robust data analysis pipelines and algorithms.
PO4	**Experimental Design and Inference:** Design, analyze, and interpret experimental results using regression, statistical inference, hypothesis testing, and advanced experimental methods.
PO5	**Machine Learning Competency:** Develop and deploy machine learning models—including supervised, unsupervised, and reinforcement learning techniques—to solve complex predictive tasks.
PO6	**Advanced Analytics and Modeling:** Use a variety of statistical and deep learning approaches (e.g., Bayesian inference, time series forecasting) to tackle domain-specific challenges.

(Continued on next page)

(Continued from previous page)

Code	Description
PO7	**Domain-Specific Applications:** Leverage knowledge of specialized fields (e.g., biostatistics, industrial statistics, business analytics) to provide data-driven insights and solutions.
PO8	**Critical Thinking and Problem-Solving:** Formulate research questions, evaluate data sources, and synthesize analytical methods to address multifaceted problems with rigor and clarity.
PO9	**Data Visualization and Communication:** Employ best practices in data visualization tools and storytelling techniques to effectively communicate findings to both technical and non-technical audiences.
PO10	**Ethics and Societal Impact:** Recognize ethical, legal, and societal considerations in data handling, algorithmic fairness, privacy, and AI regulation, ensuring responsible practice.
PO11	**Research Methodology and Innovation:** Design independent research projects employing systematic literature reviews, robust methodologies, and reproducible workflows, culminating in innovative findings.
PO12	**Collaboration and Leadership:** Demonstrate effective teamwork, leadership, and project management skills across multidisciplinary teams, aligning data science initiatives with organizational goals.
PO13	**Adaptability and Lifelong Learning:** Stay current with evolving data science trends and technologies, engaging in continuous professional development and innovative thinking.
PO14	**Internship and Practical Exposure:** Translate academic knowledge into practical solutions through hands-on internships or industry-based projects, enhancing employability and professional acumen.
PO15	**Global and Cross-Cultural Perspective:** Integrate global best practices, cross-cultural insights, and international standards in data analytics, fostering inclusive and globally relevant solutions.

Syllabus for Master of Data Science and Applied Statistics w.e.f. 2025-26

Outline of the syllabus

Semester I

Code	Course Title (Page)	Theory	Practical	Credit
DS101	Probability (p. 5)	3	1	4
DS102	Linear Algebra (p. 6)	3	1	4
DS103	Statistical Foundation (p. 7)	3	1	4
DS104	Programming using Python and R (p. 8)	0	4	4
DS105	Database Systems and SQL (p. 9)	2	2	4
PGAUC01	Elementary Arabic and Islamic Studies	4	0	0

Semester II

Code	Course Title (Page)	Theory	Practical	Credit
DS201	Regression Analysis (p. 10)	3	1	4
DS202	Statistical Inference (p. 11)	3	1	4
DS203	Multivariable Analysis and Optimization (p. 12)	3	1	4
DS204	Data Structures and Algorithms (p. 14)	3	1	4
DS205	Machine Learning (p. 15)	2	2	4
PGAEC01	Research Methodology and Ethics in AI(p. 16)/Disaster Management/Human Rights and Value Education/Yoga and Life Skills	4	0	0

Semester III

Code	Course Title (Page)	Theory	Practical	Credit
DS301	Advanced Machine Learning Techniques (p. 17)	2	2	4
DS302	Internship (p. 18)	0	4	4
DS303	Discipline Elective I (p. 18)	3	1	4
DS304	Discipline Elective II (p. 24)	3	1	4
DS305	Generic Elective I: Advanced Statistical Methods (p. 30)/Cryptography-I?Mathematical Modelling through differential equation-I	3	1	4

Semester IV

Code	Course Title (Page)	Theory	Practical	Credit
DS401	Big Data Analytics (p. 31)	2	2	4
DS402	Discipline Elective III (p. 32)	3	1	4
DS403	Discipline Elective IV (p. 39)	3	1	4
DS404	Generic Elective II: Advanced Deep Learning and AI Innovations (p. 45)	3	1	4
DS405	Project (p. 46)	0	4	4

Overall Program Credits: 4 semesters × 20 credits = 80 credits
Overall Marks (per course): 5 courses × 50 marks each semester = 250 marks/semester
Total Marks for 4 semesters: 250 × 4 = 1000 marks

Elective Type	Available Courses
Discipline Elective I	Reliability Engineering and Life Testing (p. 18),Business Analytics (p. 19), Survival Analysis (p. 20),Natural Language Processing (p. 21), Bayesian Statistics (p. 22), Environmental Data Analytics (p. 23)
Discipline Elective II	Design and Analysis of Industrial Experiments (p. 24), Time Series Analysis (p. 25), Epidemiology (p. 26), Computer Vision (p. 27), Statistical Learning Theory (p. 28), Climate Data Science/Sustainable Resource Management Analytics (p. 29)
Discipline Elective III	Quality Control and Six Sigma (p. 33), Marketing and Social Media Analytics (p. 33),Statistical Genetics and Genomics (p. 34), Reinforcement Learning (p. 35),Causal Inference and Counterfactual Analysis (p. 37), Geospatial Analytics for Environmental Planning (p. 36)
Discipline Elective IV	Supply Chain Analytics (p. 39), Financial Risk Analytics (p. 40),Design of Clinical Trials (p. 41), Cybersecurity Analytics (p. 42),Advanced Statistical Computing (p. 43), Climate Risk Modeling and Impact Assessment (p. 44)

Discipline-Specific Elective Buckets

To promote specialization and interdisciplinary integration, electives are grouped into five distinct thematic buckets:

Elective Bucket	Elective Courses
Applied Industrial Statistics	1. Reliability Engineering and Life Testing (p. 18) 2. Design and Analysis of Industrial Experiments (p. 24) 3. Quality Control and Six Sigma (p. 33) 4. Supply Chain Analytics (p. 39)
Business Intelligence	1. Business Analytics (p. 19) 2. Time Series Analysis (p. 25) 3. Marketing and Social Media Analytics (p. 33) 4. Financial Risk Analytics (p. 40)
Biostatistics	1. Survival Analysis (p. 20) 2. Epidemiology (p. 26) 3. Statistical Genetics and Genomics (p. 34) 4. Design of Clinical Trials (p. 41)
AI and Intelligent Systems	1. Natural Language Processing (p. 21) 2. Computer Vision (p. 27) 3. Reinforcement Learning (p. 35) 4. Cybersecurity Analytics (p. 42)
Core Statistical Theory and Methods	1. Bayesian Statistics (p. 22) 2. Statistical Learning Theory (p. 28) 3. Causal Inference and Counterfactual Analysis (p. 36) 4. Advanced Statistical Computing (p. 42)
Environmental Sustainability	1. Environmental Data Analytics (p. 23) 2. Climate Data Science/Sustainable Resource Management Analytics (p. 29) 3. Geospatial Analytics for Environmental Planning (p. 37) 4. Climate Risk Modeling and Impact Assessment (p. 44)

Detailed Syllabus with Key Topics and Reference Books

Semester I

DS101: Probability (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Probability	Probability definitions, probability rules, and counting techniques, Conditional Probability, Independent event, Total Probability and Bayes Theorem.	10
Random Variables	Understanding discrete and continuous random variables, PMF, PDF, and CDF.	5
Different Moments	Expected value, moments, variance, covariance, correlation, moment-generating functions, multiple random variables, and conditional expectation.	10
Important Probability Distributions	Binomial, Poisson, Negative Binomial, Uniform Normal, Log-Normal, Gamma, Beta, Cauchy, Logistic, Chi-square, t, and F-distributions, applications in statistical inference.	15
Inequality & Limit Theorems	Chebyshev and Markov Inequality, Convergence of Random Variables, WLLN, SLLN and Central Limit Theorem (CLT).	6
Applications in Data Science	Case studies on probability models in predictive analytics, risk modeling, and decision-making.	4
		End of table

Recommended Books

No.	Book Title and Author
1	*A First Course in Probability* by Sheldon Ross
2	*Probability and Statistics* by Morris H. DeGroot and Mark J. Schervish
3	*Introduction to Probability* by Dimitri P. Bertsekas and John N. Tsitsiklis
4	*Probability: Theory and Examples* by Rick Durrett
5	*All of Statistics: A Concise Course in Statistical Inference* by Larry Wasserman
	End of table

Practical Exercises

No.	Practical Description
1	Calculate basic probabilities using classical, empirical, and axiomatic approaches. Solve problems using probability rules, including addition and multiplication theorems.
2	Apply conditional probability and Bayes' Theorem to solve real-life scenarios. Simulate examples of independent and dependent events.
3	Define and plot PMFs and PDFs for discrete and continuous random variables. Use software tools to visualize distributions.
4	Compute expectation, variance, standard deviation, and higher-order moments for different distributions. Analyze relationships through covariance and correlation.
	Continued on next page

No.	Practical Description
5	Simulate and analyze common probability distributions: Binomial, Poisson, Uniform, Normal, and Exponential. Estimate and compare empirical and theoretical properties.
6	Generate and analyze samples from Gamma, Beta, Log-Normal, Cauchy, Chi-square, t, and F-distributions using Python/R.
7	Demonstrate Chebyshev's and Markov's Inequality through numerical examples. Visualize convergence concepts and simulate the Law of Large Numbers (LLN).
8	Simulate and verify the Central Limit Theorem (CLT) using different parent distributions and sample sizes.
9	Estimate unknown parameters using the method of moments and maximum likelihood estimation (MLE). Compare different estimators for bias and efficiency.
10	Conduct a mini-project or case study involving real-world data to model uncertainty and make probabilistic inferences in applications such as finance, healthcare, or reliability engineering.
	End of table

DS102: Linear Algebra (45 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Matrix Algebra and Vector Spaces	Review of matrices, operations, inverses, determinants, linear independence, basis, dimension, rank, and nullity.	10
Eigenvalues and Eigenvectors	Definitions, computation methods, diagonalization of matrices, and applications in dynamical systems and PCA.	10
Inner Product Spaces	Norms, inner products, orthogonality, orthonormal bases, Gram-Schmidt process, and projections.	10
Singular Value Decomposition (SVD)	Concept, computation, interpretation of SVD, and applications in dimensionality reduction.	5
Quadratic Forms	Real quadratic forms, matrix representation, transformations, index and signature, definiteness.	10
		End of table

Recommended Books

No.	Book Title and Author
1	*Linear Algebra and Its Applications* by Gilbert Strang
2	*Introduction to Linear Algebra* by Gilbert Strang
3	*Matrix Analysis and Applied Linear Algebra* by Carl D. Meyer
4	*No Bullshit Guide to Linear Algebra* by Ivan Savov
5	*Mathematics for Machine Learning* by Marc Peter Deisenroth, A. Faisal, and C. Ong
	End of table

Practical Exercises

No.	Practical Description
1	Compute matrix operations, inverses, and determinants using Python/R/Matlab.
2	Analyze linear dependence and compute rank and nullity of matrices.

No.	Practical Description
3	Find eigenvalues and eigenvectors of symmetric and non-symmetric matrices. Apply diagonalization.
4	Apply the Gram-Schmidt process and perform orthogonal projections in vector spaces.
5	Perform SVD on small matrices and interpret the singular values and vectors.
6	Analyze and transform quadratic forms; test for definiteness and compute canonical forms.
7	Visualize orthogonality and basis transformations using simple 2D/3D vector plots.
8	Explore real-world applications of quadratic forms such as conic sections or optimization surfaces.
	End of table

DS103: Statistical Foundation (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Population and Sampling Concepts	Concepts of population and sample. Intuitive understanding of random sampling. Sampling bias and representativeness.	5
Study Design	Observational studies vs. randomized experiments. Principles of study design and data collection methods.	5
Data Types and Collection	Types of data (qualitative/quantitative, discrete/continuous). Primary and secondary data sources.	5
Data Summarization and Visualization	Graphical representation: box plots, histograms, bar plots, pie charts, scatter plots, empirical c.d.f., Q-Q plots.	10
Descriptive Measures	Measures of location (mean, median, mode), dispersion (variance, IQR, std. dev.), skewness. Robust statistics and outlier detection.	10
Bivariate Data Analysis	Scatterplots, association measures, covariance, Pearson correlation. Introduction to simple linear regression.	10
Categorical Data Analysis	Cross-tabulation, interpretation of contingency tables, odds ratio, basic properties.	5
Sampling and Resampling Techniques	Sample generating methods (random, stratified, systematic). Resampling techniques (bootstrap, permutation tests).	5
Cross-Validation Methods	Introduction to K-Fold, Leave-One-Out (LOOCV), Stratified K-Fold for model validation.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*The Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
2	*Statistics: Unlocking the Power of Data* by Robin H. Lock et al.
3	*Introduction to the Practice of Statistics* by David S. Moore, George P. McCabe, and Bruce A. Craig
4	*R for Data Science* by Hadley Wickham and Garrett Grolemund
	End of table

Practical Exercises

No.	Practical Description
1	Data loading, inspection, and basic commands using in-built and CSV datasets in Python/R.
2	Simulate and visualize random samples from different distributions. Explore sampling variability.
3	Construct and interpret histograms, box plots, bar charts, and Q-Q plots using R's ggplot2 package.
4	Calculate descriptive statistics (mean, median, standard deviation, IQR) and identify outliers.
5	Compare robust vs. non-robust measures on skewed datasets.
6	Explore bivariate data using scatterplots. Compute and interpret Pearson correlation coefficients.
7	Fit a simple linear regression model. Interpret slope, intercept, and residual plots.
8	Create and analyze cross-tabulations. Compute and interpret odds ratios.
9	Apply bootstrapping and permutation tests using R. Visualize resampling distributions.
10	Implement and compare K-Fold, LOOCV, and Stratified K-Fold cross-validation using the 'caret' or 'rsample' package.
	End of table

DS104: Programming using Python and R (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Programming	Basics of Python/R, syntax, data types, and control structures.	5
Functions and Object-Oriented Programming	Writing functions, classes, and understanding OOP concepts.	10
Data Structures	Lists, tuples, dictionaries, and sets in Python/R.	5
File Handling and APIs	Reading/writing files, exception handling, and API integration.	5
Data Visualization Basics	Introduction to Matplotlib, Seaborn, and Plotly.	5
Advanced Visualization Techniques	Interactive visualizations, time-series analysis, and geospatial visualizations.	10
Dashboard Creation	Building dashboards with Tableau and Power BI.	5
NumPy and Pandas for Data Manipulation	Handling and analyzing data efficiently.	5
Regular Expressions and Functional Programming	Text processing and advanced programming techniques.	5
Web Scraping and Parallel Computing	Collecting data from websites and basic multi-threading concepts.	5
Debugging, Testing, and Machine Learning Applications	Writing efficient code, unit testing, and introduction to ML applications.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Python for Data Analysis* by Wes McKinney
2	*R for Data Science* by Hadley Wickham and Garrett Grolemund
3	*Data Visualization: A Practical Introduction* by Kieran Healy
4	*Python Data Science Handbook* by Jake VanderPlas
5	*Interactive Data Visualization for the Web* by Scott Murray
	End of table

Practical Exercises

No.	Practical Description
1	Write basic Python and R programs using variables, data types, input/output, loops, and conditionals.
2	Define and use custom functions and classes; demonstrate object-oriented principles like inheritance and encapsulation.
3	Work with lists, tuples, dictionaries, and sets. Apply indexing, slicing, and built-in functions for data manipulation.
4	Perform file operations (read/write text and CSV files), handle exceptions, and fetch data from APIs (e.g., RESTful JSON).
5	Create basic plots using Matplotlib, Seaborn, and ggplot2 in R. Visualize trends and distributions.
6	Develop advanced visualizations including animated, time-series, and geographic plots using Plotly and GeoPandas.
7	Build interactive dashboards using Tableau and Power BI with real datasets. Connect data sources and apply filters.
8	Use NumPy arrays and Pandas DataFrames for data wrangling, cleaning, aggregation, and reshaping.
9	Apply regular expressions for pattern matching in text data. Demonstrate use of map, filter, and lambda functions.
10	Scrape data from websites using ‘requests‘ and ‘BeautifulSoup‘ or ‘rvest‘. Implement basic multithreading for parallel tasks.
11	Use debugging tools, write unit tests (e.g., pytest/unittest), and implement a small ML pipeline (e.g., linear regression).
	End of table

DS105: Database Systems and SQL (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Databases	Database concepts, DBMS vs. file system, architecture of DBMS, data independence, real-life applications.	5
Data Models and ER Diagrams	Entity-Relationship (ER) model, attributes, keys, relationship sets, ER-to-relational mapping.	5
Relational Model	Relational schema, domains, tuples, relational integrity constraints (key, domain, referential).	5
		Continued on next page...

Topic	Description	Hours
Relational Algebra and Calculus	Selection, projection, joins, union, intersection, set difference, division; tuple and domain relational calculus.	5
SQL Basics	Introduction to SQL, data types, DDL, DML, simple queries using SELECT, INSERT, UPDATE, DELETE.	5
Advanced SQL Queries	Joins (INNER, OUTER, SELF), subqueries, GROUP BY, HAVING, views, set operations, aggregate functions.	10
Constraints and Normalization	Primary key, foreign key, unique and check constraints; functional dependencies, 1NF, 2NF, 3NF, BCNF.	5
Transaction Management	ACID properties, transactions, concurrency control, schedules, serializability, recovery mechanisms.	5
Case Studies and Projects	Hands-on database design, query writing, indexing, and optimization tasks on real-world datasets.	2
		End of table

Recommended Books

No.	Book Title and Author
1	*Database System Concepts* by Abraham Silberschatz, Henry Korth, and S. Sudarshan
2	*Fundamentals of Database Systems* by Ramez Elmasri and Shamkant B. Navathe
3	*SQL for Data Scientists* by Renee M. P. Teate
4	*Seven Databases in Seven Weeks* by Eric Redmond and Jim R. Wilson
5	*NoSQL Distilled* by Pramod J. Sadalage and Martin Fowler
	End of table

Practical Exercises

No.	Practical Description
1	Design and create ER diagrams for given case studies and convert them to relational schema.
2	Implement relational databases using MySQL/PostgreSQL.
3	Write and execute basic SQL queries including SELECT, INSERT, UPDATE, DELETE.
4	Use joins and subqueries to extract meaningful insights from datasets.
5	Create and manipulate views, use aggregate functions with GROUP BY and HAVING.
6	Apply normalization techniques on sample datasets and decompose into normalized forms.
7	Simulate transaction scenarios and analyze concurrent transactions and conflicts.
8	Create and evaluate indexes, analyze query plans, and optimize queries.
9	Work with a NoSQL database (MongoDB or Firebase), store and query JSON-like data.
10	Mini-project: Design a full-stack database-driven application with schema design, complex queries, and optimization.
	End of table

Semester II

DS201: Regression Analysis (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Simple Linear Regression	Model formulation, least squares estimation, and interpretation.	5
Multiple Linear Regression	Assumptions, diagnostics, and variance inflation factor (VIF) for multicollinearity detection.	10
Logistic and Polynomial Regression	Binary classification using logistic regression and higher-order polynomial regression.	10
ANOVA and ANCOVA	One-way and two-way ANOVA, interaction effects, and combining with regression in ANCOVA.	10
Generalized Linear Models (GLM)	Introduction to Poisson regression, logistic regression extensions, and applications.	5
Model Selection Techniques	AIC, BIC, cross-validation, and best subset selection methods.	5
Residual Analysis and Transformations	Checking model assumptions and improving model fit with transformations.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Applied Regression Analysis* by Norman Draper and Harry Smith
2	*Regression Modeling Strategies* by Frank Harrell
3	*The Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
4	*An Introduction to Generalized Linear Models* by Annette J. Dobson and Adrian G. Barnett
	End of table

Practical Exercises

No.	Practical Description
1	Implement simple linear regression and interpret coefficients using real-world datasets.
2	Conduct multiple linear regression and diagnose multicollinearity using VIF.
3	Fit and evaluate logistic and polynomial regression models.
4	Perform one-way and two-way ANOVA and ANCOVA, and interpret results.
5	Apply model selection techniques using AIC, BIC, and cross-validation.
6	Implement Ridge, Lasso, and Elastic Net regularization methods.
7	Use GLM (e.g., Poisson regression) for modeling count data.
8	Conduct residual analysis and apply data transformations to improve model assumptions.
	End of table

DS202: Statistical Inference (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Estimation	Overview of point estimation, properties of estimators, Maximum Likelihood Estimation.	10
Introduction to Hypothesis Testing	Formulation of null and alternative hypotheses, Type I and Type II errors, power function, Neyman-Pearson Lemma, applications in Uniformly Most Powerful (UMP) test construction.	10
Likelihood Ratio Tests	Concept, applications, and comparison with traditional methods.	5
Parametric Hypothesis Testing	Z-test, t-tests, F-test, and chi-square tests for independence and goodness-of-fit.	10
Confidence Intervals	Construction of confidence intervals for population parameters.	5
Non-Parametric Tests	Wilcoxon, Mann-Whitney U, Kruskal-Wallis, and other distribution-free methods.	5
Bayesian Inference and Hypothesis Testing	Priors, likelihoods, posteriors, Bayes factors, posterior odds, and decision-making.	10
Case Studies and Model Evaluation	Real-world applications of inference.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Statistical Inference* by George Casella and Roger L. Berger
2	*All of Statistics* by Larry Wasserman
3	*Testing Statistical Hypotheses* by Erich Lehmann and Joseph Romano
4	*Introduction to the Theory of Statistical Inference* by Hannelore Liero and Silvelyn Zwanzig
	End of table

Practical Exercises

No.	Practical Description
1	Simulate hypothesis testing scenarios and analyze Type I and Type II errors. Apply the Neyman-Pearson Lemma to construct UMP tests.
2	Perform likelihood ratio tests and compare results with traditional tests.
3	Apply Z-test, t-tests, F-test, and chi-square tests on real datasets.
4	Construct and interpret confidence intervals for mean, variance, and proportion.
5	Conduct non-parametric tests such as Wilcoxon, Mann-Whitney U, and Kruskal-Wallis.
6	Implement Bayesian inference using conjugate priors and posterior distributions. Conduct Bayesian hypothesis testing.
7	Perform bootstrap and permutation tests to estimate variability and validate inference results.
8	Apply FDR control techniques including Bonferroni and Benjamini-Hochberg procedures.
9	Design and evaluate A/B testing experiments for digital or healthcare applications.
10	Perform full-case analysis on a dataset involving hypothesis testing, model selection, and interpretation of inference results.
	End of table

DS203: Multivariable Analysis and Optimization (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Functions of Several Variables	Introduction to multivariable functions, graphical representation, limits and continuity in higher dimensions.	5
Gradients and Partial Derivatives	Partial derivatives, gradient vectors, directional derivatives, and applications in optimization and sensitivity analysis.	5
Jacobians and Hessians	Computation of Jacobians and Hessians, interpretation, role in multivariable change of variables, convexity analysis, and optimization contexts.	5
Optimization Techniques	Constrained and unconstrained optimization, Lagrange multipliers, and applications to real-world problems.	5
Advanced Optimization Algorithms	Introduction to numerical optimization algorithms: Stochastic Gradient Descent (SGD), BFGS, convergence criteria, and implementation issues.	5
Multiple Integration	Double and triple integrals, iterated integrals, change of order, and applications in volume and probability computations.	5
Line and Surface Integrals	Definitions, evaluation techniques, applications in physics and probability, Green's and Stokes' Theorems (intuitive).	5
Applications in Data Science	Case studies from machine learning, optimization in model training, gradient-based learning, and real-world engineering/data problems.	10

End of table

Recommended Books

No.	Book Title and Author
1	*Advanced Engineering Mathematics* by Erwin Kreyszig
2	*Multivariable Calculus* by James Stewart
3	*Convex Optimization* by Stephen Boyd and Lieven Vandenberghe
4	*Mathematics for Machine Learning* by Marc Peter Deisenroth, A. Faisal, and C. Ong
5	*Numerical Methods for Engineers* by Steven C. Chapra and Raymond P. Canale

End of table

Practical Exercises

No.	Practical Description
1	Compute gradients and directional derivatives of multivariable functions using symbolic and numerical tools.
2	Evaluate Jacobian and Hessian matrices and analyze convexity of functions.
3	Apply Lagrange multipliers to solve constrained optimization problems.
4	Implement gradient descent and stochastic gradient descent for simple regression problems.
5	Use BFGS and other quasi-Newton methods with Python or R optimization libraries.
6	Compute double and triple integrals both analytically and numerically.
7	Evaluate line and surface integrals and visualize vector fields.
8	Solve ordinary differential equations using analytical and numerical methods.
9	Implement numerical methods for integration and root finding in Python.

Continued on next page

No.	Practical Description
10	Case study: Train a logistic regression model using gradient-based optimization and analyze convergence.
	End of table

DS204: Data Structures & Algorithms (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Algorithm Complexity Analysis	Big-O, Big-Theta, and Big-Omega notations, time and space complexity.	5
Arrays and Linked Lists	Introduction, operations, advantages, and disadvantages.	5
Stacks and Queues	Implementation, applications, and real-world use cases.	5
Trees and Binary Search Trees	Properties, traversal techniques, and applications.	10
Graph Representations and Traversals	BFS, DFS, adjacency lists, and adjacency matrices.	10
Sorting Algorithms	Merge Sort, Quick Sort, Heap Sort, and their complexities.	10
Searching Algorithms	Binary Search, Hashing, and their implementations.	5
Advanced Algorithmic Techniques	Greedy algorithms, dynamic programming, recursion, and backtracking.	5
Shortest Path Algorithms	Dijkstra's and Bellman-Ford algorithms and their applications.	5
Applications in Machine Learning	Data structures used in ML, optimization techniques, and real-world case studies.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Introduction to Algorithms* by Cormen, Leiserson, Rivest, and Stein
2	*Algorithms* by Robert Sedgewick and Kevin Wayne
3	*Data Structures and Algorithm Analysis in C++* by Mark Allen Weiss
4	*The Algorithm Design Manual* by Steven S. Skiena
	End of table

Practical Exercises

No.	Practical Description
1	Implementing arrays and linked lists with basic operations.
2	Implementing stacks and queues using arrays and linked lists.
3	Constructing and traversing binary search trees (inorder, preorder, postorder).
4	Implementing graph representations and BFS/DFS traversals.
5	Writing and analyzing sorting algorithms (Merge Sort, Quick Sort, Heap Sort).
6	Implementing searching algorithms (Binary Search, Hashing).
	Continued on next page...

No.	Practical Description
7	Solving problems using dynamic programming and recursion (e.g., Fibonacci, Knapsack).
8	Implementing shortest path algorithms (Dijkstra's and Bellman-Ford).
9	Implementing string matching algorithms (KMP, Rabin-Karp, Boyer-Moore).
10	Applying data structures in machine learning (e.g., decision trees, priority queues, hash maps).
	End of table

DS205: Machine Learning (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Machine Learning	Definition, importance, and real-world applications of ML. Overview of learning paradigms: supervised, unsupervised, semi-supervised, reinforcement learning.	5
Bias-Variance Tradeoff	Understanding overfitting, underfitting, and generalization.	5
Data Preprocessing and Feature Engineering	Data cleaning, handling missing values/outliers, feature scaling (normalization, standardization), encoding categorical variables (one-hot, label encoding).	10
Dimensionality Reduction	Principal Component Analysis (PCA), t-SNE, and Linear Discriminant Analysis (LDA).	5
Supervised Learning Algorithms	Review of Regression Techniques and LDA, QDA, Naive Bayes KNN, idea of PCR, Decision Trees, Random Forests, Support Vector Machines (SVM).	10
Model Evaluation	Accuracy, precision, recall, F1-score, confusion matrix, ROC Curve, AUC Score.	5
Unsupervised Learning Algorithms	Clustering: K-Means, Hierarchical Clustering, Gaussian Mixture Models (GMM); Association rule learning: Apriori, FP-Growth.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Pattern Recognition and Machine Learning* by Christopher M. Bishop
2	*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* by Aurélien Géron
3	*The Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
4	*Machine Learning: A Probabilistic Perspective* by Kevin P. Murphy
	End of table

Practical Exercises

No.	Practical Description
1	Implementing data preprocessing techniques: handling missing values, feature scaling, and encoding categorical variables.
	Continued on next page

No.	Practical Description
2	Applying Principal Component Analysis (PCA) and t-SNE for dimensionality reduction and visualization.
3	Implementing Linear Regression, Polynomial Regression, and Logistic Regression models.
4	Training and evaluating Decision Trees, Random Forests, and Support Vector Machines (SVM).
5	Applying ensemble learning methods: Bagging, Boosting (XGBoost, AdaBoost), and Stacking.
6	Evaluating models using cross-validation and hyperparameter tuning (Grid Search, Random Search).
7	Implementing clustering algorithms: K-Means, Hierarchical Clustering, DBSCAN, and Gaussian Mixtures.
8	Performing association rule learning with Apriori and FP-Growth algorithms.
9	Building and training Multi-layer Perceptrons (MLPs) with ReLU, Sigmoid, and Softmax activation functions.
10	Optimizing neural networks using SGD, Adam, and RMSprop; evaluating training performance.
	End of table

PGAEC01: Research Methodology and Ethics in AI (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Research Methodology	Types of research, research process, formulating research problems and hypotheses, and literature review techniques.	5
Research Design and Data Collection	Qualitative vs. quantitative approaches, sampling methods, experimental design, and ethical data collection practices.	5
Statistical Tools for Research	Basics of statistical inference, regression, hypothesis testing, and using statistical software in research.	5
Introduction to AI Ethics	Overview of ethical concerns in AI, social responsibility, and impact of AI on society.	5
Fairness and Accountability in AI	Fairness definitions, algorithmic bias, explainability, transparency, and frameworks like GDPR and AI Act.	10
Privacy, Security, and Deepfakes	Differential privacy, federated learning, adversarial attacks, deepfake generation/detection, and misinformation.	10
Research Integrity and Publication Ethics	Plagiarism, authorship, peer review, conflict of interest, and ethical considerations in publishing.	5
AI for Social Good and Case Studies	Use of AI in healthcare, education, and environment with ethical implications. Analysis of real-world bias and regulation cases.	10
		End of table

No.	Book Title and Author
1	*The Ethical Algorithm: The Science of Socially Aware Algorithm Design* by Michael Kearns and Aaron Roth
2	*Weapons of Math Destruction* by Cathy O'Neil
3	*Fairness and Machine Learning: Limitations and Opportunities* by Solon Barocas, Moritz Hardt, and Arvind Narayanan
4	*Research Methodology: A Step-by-Step Guide for Beginners* by Ranjit Kumar
5	*Responsible AI: Best Practices for Creating Trustworthy AI Systems* by Virginia Dignum

Semester III

DS301: Advanced Machine Learning Techniques (60 Lectures) (Back to Table)

Course Content

Topic	**Description**	**Hours**
Advanced Supervised Learning	Ridge Regression, Lasso Regression, Bayesian Regression, and advanced decision trees (XGBoost, CatBoost, LightGBM).	10
Hyperparameter Tuning	Bayesian Optimization, Genetic Algorithms, and tuning strategies for complex models.	5
Advanced Unsupervised Learning	Spectral Clustering, Mean Shift, Topic Modeling (LDA), UMAP, and Autoencoders.	10
Deep Learning Fundamentals	CNNs for image processing, Transfer Learning (VGG, ResNet, EfficientNet), RNNs, LSTMs, GRUs, and Transformers (BERT, GPT).	10
Model Explainability	Feature importance, SHAP values, LIME, fairness, bias detection, and ethical AI considerations.	5
Real-World Applications	Case studies in computer vision, NLP, healthcare, finance, and robotics.	5
		End of table

Recommended Books

No.	**Book Title and Author**
1	*Machine Learning: A Probabilistic Perspective* by Kevin P. Murphy
2	*Ensemble Methods in Machine Learning* by Zhi-Hua Zhou
3	*Bayesian Reasoning and Machine Learning* by David Barber
4	*Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
	End of table

Practical Exercises

No.	**Practical Description**
1	Implement Ridge, Lasso, and Bayesian Regression models; evaluate regularization effects.
2	Train and tune advanced decision trees: XGBoost, CatBoost, and LightGBM. Analyze performance.
3	Perform high-dimensional clustering using Spectral Clustering and Mean Shift algorithms.
4	Apply Latent Dirichlet Allocation (LDA) for unsupervised topic modeling on text datasets.
5	Implement MDPs and Q-Learning; simulate simple grid-world environments.
6	Train Deep Q-Networks (DQN) for basic reinforcement learning problems using OpenAI Gym.
7	Build CNNs for image classification using custom and standard datasets (e.g., MNIST, CIFAR-10).
8	Use transfer learning with pre-trained models like VGG16, ResNet50, and EfficientNet for fine-tuning.
9	Train LSTM and GRU networks for sequential data (time series and NLP tasks).
	Continued on next page...

No.	Practical Description
10	Apply model explainability techniques: SHAP, LIME; assess model bias and fairness on structured data.
	End of table

DS302: Internship (Back to Table)

Internship Evaluation Criteria (After Semester II, Before Semester III)

According to Aliah University Internship Criteria
DS303: Discipline Elective I (Choose one) (Back to Table)

Students must choose one elective from the following options:
DS303A: Reliability Engineering and Life Testing (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Reliability	Concepts of reliability, failure, and maintainability; bathtub curve	5
Lifetime Distributions	Exponential, Weibull, Gamma and Lognormal distributions	10
System Reliability Models	Series, parallel, and k-out-of-n systems, minimal path and cut sets	10
Estimation of Reliability Parameters	MLE, confidence intervals for reliability functions, Bayesian approach	10
Accelerated Life Testing	Principles and models for accelerated tests, degradation testing	5
Reliability Growth Models	Duane model, AMSAA model, Crow model for repairable systems	5
Maintenance Models	Preventive and corrective maintenance, availability modeling	5
Life Data Analysis	Graphical and statistical methods, goodness-of-fit tests	5
Applications and Case Studies	Real-world applications in engineering and manufacturing	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Reliability Engineering* by Elsayed A. Elsayed
2	*Statistical Methods for Reliability Data* by William Q. Meeker and Luis A. Escobar
3	*Practical Reliability Engineering* by Patrick O'Connor and Andre Kleyner
4	*Life Data Analysis* by Wayne B. Nelson
	End of table

Practical Exercises

No.	Practical Description
1	Fit lifetime distributions to failure data using MLE methods.
2	Compute system reliability for series and parallel configurations.
	Continued on next page

No.	Practical Description
3	Analyze accelerated life test data using graphical and statistical methods.
4	Estimate reliability parameters and confidence intervals.
5	Apply reliability growth models to repairable systems.
6	Conduct preventive and corrective maintenance modeling.
7	Perform life data analysis using real-world datasets.
8	Simulate failure data and conduct reliability prediction.
9	Use software tools (e.g., R, MATLAB, Minitab) for reliability computations.
10	Case study: Evaluate and report on a reliability improvement project.
	End of table

DS303B: Business Analytics (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Business Analytics	Overview of data-driven decision-making and key business metrics.	5
Market Segmentation	Clustering techniques (K-Means, Hierarchical Clustering) for customer segmentation.	10
Customer Profiling	Identifying customer behaviors and lifetime value (RFM analysis, cohort analysis).	10
Predictive Modeling for Business	Regression models, classification techniques, and machine learning in business forecasting.	10
Business Intelligence Tools	Dashboard creation using Tableau, Power BI, and SQL-based analytics.	10
Optimization for Decision Making	Linear programming, A/B testing, and decision trees for strategic planning.	5
Text Analytics in Business	Sentiment analysis, NLP for customer feedback, and market trends.	5
Real-World Applications	Case studies in retail, e-commerce, finance, and supply chain analytics.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Competing on Analytics: The New Science of Winning* by Thomas H. Davenport and Jeanne G. Harris
2	*Data Science for Business* by Foster Provost and Tom Fawcett
3	*Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* by Eric Siegel
4	*Business Analytics: Data Analysis & Decision Making* by S. Christian Albright and Wayne L. Winston
	End of table

List of Numerical Practicals

Practical Exercises

No.	Practical Description
1	Performing exploratory data analysis (EDA) on business datasets.
2	Implementing clustering techniques for market segmentation.
3	Building predictive models for customer churn analysis.
4	Creating customer profiles using RFM analysis.
5	Conducting sentiment analysis on customer reviews.
6	Designing interactive dashboards using Tableau and Power BI.
7	Running A/B testing experiments for business decision-making.
8	Forecasting sales using regression models.
9	Optimizing business strategies using linear programming.
10	Case study: Business analytics in e-commerce and digital marketing.
	End of table

DS303C: Survival Analysis (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Survival Analysis	Basic concepts: time-to-event data, censoring, survival function, hazard function. Applications in healthcare, engineering, and social sciences.	5
Non-parametric Methods	Kaplan-Meier estimator, life tables, log-rank test for comparing survival curves.	10
Parametric Survival Models	Exponential, Weibull, and other parametric models. Estimation of survival and hazard functions.	10
Semi-parametric Models: Cox Proportional Hazards Model	Cox model formulation, proportional hazards assumption, partial likelihood estimation.	10
Model Diagnostics	Assessing model fit, testing proportional hazards assumption, Schoenfeld residuals, and influence diagnostics.	10
Time-varying Covariates	Modeling time-dependent predictors and interactions in survival models.	5
Competing Risks and Recurrent Events	Analysis when subjects are exposed to multiple event types or recurring events.	5
Applications and Case Studies	Real-life datasets, interpretation of results, and reporting in survival analysis.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Survival Analysis: A Self-Learning Text* by David G. Kleinbaum and Mitchel Klein
2	*Applied Survival Analysis: Regression Modeling of Time-to-Event Data* by David W. Hosmer, Stanley Lemeshow, and Susanne May
3	*Modelling Survival Data in Medical Research* by David Collett
4	*The Statistical Analysis of Failure Time Data* by John D. Kalbfleisch and Ross L. Prentice
	Continued on next page

No.	Book Title and Author
5	*Survival Analysis Using SAS: A Practical Guide* by Paul D. Allison
	End of table

List of Numerical Practicals

Numerical Practicals for Survival Analysis

No.	Practical Description
1	Estimating survival functions using the Kaplan-Meier method.
2	Comparing survival curves using the log-rank test.
3	Fitting and interpreting exponential and Weibull survival models.
4	Building and validating Cox proportional hazards models.
5	Conducting residual analysis for model diagnostics (e.g., Schoenfeld residuals).
6	Handling time-varying covariates in Cox models.
7	Analyzing datasets with censoring and truncation.
8	Survival analysis with competing risks.
9	Visualizing survival data using survival curves and hazard plots.
10	Case study: Survival analysis in a clinical trial or customer churn scenario.
	End of table

DS303D: Natural Language Processing (60 Lectures) (Back to Table)

Course Content - NLP

Topic	Description	Hours
Introduction to NLP	Overview of NLP, text preprocessing techniques, tokenization, stemming, lemmatization, and stopword removal.	5
Word Embeddings	Understanding word embeddings: Word2Vec, GloVe, and FastText.	10
Text Classification	Sentiment analysis, spam detection, and text categorization using ML models.	5
Named Entity Recognition (NER)	Identifying entities in text using rule-based, statistical, and deep learning methods.	5
Topic Modeling	Implementing Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) for document clustering.	5
Sequence Modeling	Applications of Recurrent Neural Networks (RNNs), LSTMs, and GRUs in NLP tasks.	10
Transformers in NLP	Introduction to transformer models such as BERT, GPT, and T5 for text generation and classification.	10
Text Summarization	Abstractive and extractive text summarization using deep learning models.	5
Machine Translation	Sequence-to-sequence models and attention mechanisms for language translation.	5
Real-World Applications	Case studies on NLP applications in healthcare, finance, and customer service.	5
		End of table

Reference Books (NLP)

Recommended Books - NLP

No.	Book Title and Author
1	*Speech and Language Processing* by Daniel Jurafsky and James H. Martin
2	*Natural Language Processing with Python* by Steven Bird, Ewan Klein, and Edward Loper
3	*Deep Learning for NLP* by Palash Goyal, Sumit Pandey, and Karan Jain
4	*Transformers for NLP* by Denis Rothman
	End of table

List of Numerical Practicals (NLP)

Practical Exercises - NLP

No.	Practical Description
1	Text preprocessing: tokenization, stemming, lemmatization, and stopword removal.
2	Implementing word embeddings using Word2Vec and GloVe.
3	Building a sentiment analysis classifier using Naïve Bayes and deep learning.
4	Performing Named Entity Recognition (NER) with spaCy and BERT.
5	Applying topic modeling with LDA and NMF.
6	Training RNNs, LSTMs, and GRUs for sequence modeling tasks.
7	Implementing BERT for text classification and question answering.
8	Extractive and abstractive text summarization using deep learning.
9	Machine translation using sequence-to-sequence models and attention mechanisms.
10	Case study: Real-world NLP application in customer support automation.
	End of table

DS303E: Bayesian Statistics (60 Lectures) (Back to Table)

Course Content - Bayesian Statistics

Topic	Description	Hours
Introduction to Bayesian Statistics	Overview of Bayesian probability, Bayes' theorem, and prior-posterior updates.	5
Bayesian Inference	Concepts of prior, likelihood, posterior, and conjugate priors.	10
Markov Chain Monte Carlo (MCMC) Methods	Gibbs Sampling, Metropolis-Hastings algorithm, and Hamiltonian Monte Carlo.	10
Bayesian Regression	Bayesian linear regression, Bayesian logistic regression, and comparison with frequentist approaches.	10
Bayesian Hierarchical Models	Multilevel models, applications in hierarchical data, and shrinkage estimators.	5
Bayesian Model Selection	Bayes factors, Bayesian Information Criterion (BIC), and posterior predictive checks.	5
Approximate Bayesian Computation (ABC)	Methods for inference when likelihoods are intractable.	5
Applications in Machine Learning	Bayesian optimization, Gaussian processes, and variational inference.	5

Continued on next page...

Topic	Description	Hours
Real-World Applications	Case studies in healthcare, finance, and ecological modeling.	5
		End of table

Reference Books (Bayesian Statistics)

Recommended Books - Bayesian Statistics

No.	Book Title and Author
1	*Bayesian Data Analysis* by Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin
2	*Bayesian Statistics the Fun Way* by Will Kurt
3	*Doing Bayesian Data Analysis* by John K. Kruschke
4	*The Bayesian Choice* by Christian Robert
	End of table

List of Numerical Practicals (Bayesian Statistics)

Practical Exercises - Bayesian Statistics

No.	Practical Description
1	Implementing Bayes' theorem for simple probabilistic inference.
2	Computing posterior distributions for different prior-likelihood combinations.
3	Simulating MCMC methods using Gibbs Sampling.
4	Implementing Metropolis-Hastings for Bayesian parameter estimation.
5	Performing Bayesian linear and logistic regression using PyMC3 or Stan.
6	Constructing Bayesian hierarchical models for multilevel data.
7	Bayesian model comparison using Bayes factors and BIC.
8	Applying Approximate Bayesian Computation (ABC) for likelihood-free inference.
9	Implementing Gaussian processes for Bayesian optimization.
10	Case study: Bayesian analysis of financial risk modeling.
	End of table

DS303F: Environmental Data Analytics (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Environmental Datasets	Types of environmental data: air, water, soil, climate, ecological; spatio-temporal nature and sources.	5
Data Cleaning and Preprocessing	Handling missing values, outlier detection, time-alignment, unit harmonization, regridding techniques.	10
Exploratory Data Analysis	Distributional analysis, trend detection, correlation matrices, time series plots, spatial heatmaps.	10
Environmental Statistical Modeling	Regression, time series decomposition, seasonality, anomaly detection.	10
		Continued on next page

Topic	Description	Hours
Machine Learning Applications	Supervised/unsupervised ML for pollutant classification, forecasting, and pattern recognition.	10
Case Studies	Case studies on air quality, water contamination, ecosystem degradation.	5
Environmental Policy Analytics	Data-driven decision support for emissions control, waste management, and public health.	10
		End of table

DS304: Discipline Elective II (Choose one) (Back to Table)

Students must choose one elective from the following options:

DS304A: Design and Analysis of Industrial Experiments (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Experimental Design	Basic principles: randomization, replication, blocking. Role of design in industrial settings.	5
Completely Randomized and Randomized Block Designs	CRD, RBD – structure, analysis, assumptions, and efficiency.	10
Latin Square and Graeco-Latin Designs	Blocking in two directions, model structure and interpretation.	5
Factorial Designs	2^k factorial designs, confounding, interaction effects, and resolution.	10
Fractional Factorial Designs	Concepts of aliasing, resolution, and screening experiments.	10
Response Surface Methodology	First and second-order designs, optimization using contour and surface plots.	10
Taguchi Methods	Robust design, signal-to-noise ratio, orthogonal arrays.	5
DOE Software and Industrial Case Studies	Use of R, Minitab, or JMP for design and analysis; real-world examples.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Design and Analysis of Experiments* by Douglas C. Montgomery
2	*Statistical Design* by George E. P. Box, William G. Hunter, and J. Stuart Hunter
3	*Industrial Experimentation* by D.C. Montgomery and G.C. Runger
4	*Taguchi Techniques for Quality Engineering* by Philip J. Ross
	End of table

Practical Exercises

No.	Practical Description
1	Conduct CRD and RBD with appropriate blocking.
2	Analyze factorial designs and interpret main and interaction effects.
3	Construct and analyze fractional factorial designs.
4	Implement response surface methodology to optimize processes.
5	Apply Taguchi design principles using orthogonal arrays.
6	Use statistical software to design and analyze industrial experiments.
7	Explore industrial case studies and replicate DOE approaches.
8	Validate assumptions and analyze residuals for model adequacy.
9	Visualize design results using contour and interaction plots.
10	Present design of experiments report with industrial recommendations.
	End of table

DS304B: Time Series Analysis (60 Lectures) (Back to Table)

Topic	Description	Hours
Introduction to Time Series	Overview of time series data, components (trend, seasonality, cycle, noise), and real-world applications.	5
Stationarity and Differencing	Understanding stationarity, unit root tests (ADF, KPSS), and differencing techniques.	5
Autoregressive (AR) and Moving Average (MA) Models	Definition, estimation, and properties of AR and MA models.	10
ARMA and ARIMA Models	Model identification, parameter estimation, diagnostic checking, and forecasting.	10
Seasonal Time Series Models	Seasonal decomposition, SARIMA models, and exponential smoothing techniques.	5
GARCH Models	Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models for volatility forecasting in financial applications.	5
Multivariate Time Series Analysis	Vector Autoregression (VAR), Vector Error Correction Models (VECM), and cointegration tests.	5
Deep Learning for Time Series	Introduction to Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks.	5
Time Series Model Evaluation	Accuracy metrics (RMSE, MAPE, MAE), cross-validation techniques, and residual analysis.	5
Anomaly Detection in Time Series	Techniques for detecting anomalies using statistical and machine learning approaches.	5
Real-World Applications	Case studies in finance, weather prediction, healthcare, and sales forecasting.	5

No.	Book Title and Author
1	*Time Series Analysis and Its Applications* by Robert H. Shumway, David S. Stoffer
2	*Forecasting: Principles and Practice* by Rob J. Hyndman, George Athanasopoulos
3	*The Analysis of Time Series: An Introduction* by Chris Chatfield
4	*Time Series Analysis* by James D. Hamilton

List of Numerical Practicals

No.	Practical Description
1	Exploratory data analysis (EDA) of time series datasets.
2	Checking stationarity using ADF and KPSS tests, performing differencing.
3	Implementing AR and MA models for time series forecasting.
4	Building ARIMA and SARIMA models for trend and seasonality forecasting.
5	Applying GARCH models for volatility estimation in financial data.
6	Implementing Vector Autoregression (VAR) for multivariate time series.
7	Training and evaluating LSTM models for sequence forecasting.
8	Performing model evaluation using RMSE, MAPE, and residual diagnostics.
9	Anomaly detection in time series using machine learning techniques.
10	Case studies on real-world forecasting applications in finance and weather prediction.

DS304C: Epidemiology (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Epidemiology	History, scope, and uses of epidemiology, key concepts: incidence, prevalence, risk, rate, odds.	5
Study Designs in Epidemiology	Descriptive, analytical, and experimental studies; cohort, case-control, and cross-sectional designs.	10
Measures of Disease Frequency and Association	Risk ratios, odds ratios, rate ratios, prevalence, incidence, and standardization methods.	10
Sources and Bias in Epidemiologic Studies	Confounding, selection bias, information bias, strategies for control in design and analysis.	10
Outbreak Investigation and Surveillance	Steps in outbreak investigation, surveillance systems, syndromic surveillance, reporting systems.	10
Screening and Diagnostic Tests	Sensitivity, specificity, predictive values, ROC curves, likelihood ratios, and applications.	5
Causal Inference in Epidemiology	Hill's criteria, counterfactual model, association vs. causation, epidemiologic triad.	5
Modern Applications and Case Studies	Applications in infectious disease epidemiology, non-communicable diseases, environmental and genetic epidemiology.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Epidemiology: An Introduction* by Kenneth J. Rothman
2	*Modern Epidemiology* by Rothman, Greenland, and Lash
3	*Essentials of Epidemiology in Public Health* by Ann Aschengrau and George R. Seage
4	*Principles of Epidemiology* by Centers for Disease Control and Prevention (CDC)
5	*Epidemiology: Beyond the Basics* by Moyses Szklo and F. Javier Nieto
	End of table

List of Numerical Practicals

Numerical Practicals

No.	Practical Description
1	Calculation of incidence and prevalence from survey data.
2	Estimating relative risk and odds ratio in cohort and case-control studies.
3	Adjusting rates using direct and indirect standardization methods.
4	Constructing and interpreting ROC curves for screening test evaluation.
5	Computing sensitivity, specificity, PPV, NPV from test results.
6	Simulating outbreak data and applying steps for outbreak investigation.
7	Identifying and adjusting for confounding using stratification and regression.
8	Designing and analyzing a case-control study from given data.
9	Conducting time-trend analysis in disease surveillance datasets.
10	Case study: Epidemiological analysis of COVID-19 data across regions.
	End of table

DS304D: Computer Vision (60 Lectures) (Back to Table)

Course Content - Computer Vision

Topic	Description	Hours
Introduction to Computer Vision	Overview of image processing, applications, and history of computer vision.	5
Image Processing Fundamentals	Image formation, color models, filtering, edge detection, and feature extraction.	10
Convolutional Neural Networks (CNNs)	CNN architecture, layers (convolution, pooling), backpropagation, and optimization.	10
Object Detection	Object localization, Faster R-CNN, YOLO, SSD, and their applications.	10
Image Segmentation	Semantic segmentation, instance segmentation, and Mask R-CNN.	5
Transfer Learning in Vision	Using pretrained models (VGG, ResNet, EfficientNet) for vision tasks.	5
Generative Models	Autoencoders, Variational Autoencoders (VAEs), and Generative Adversarial Networks (GANs).	5
3D Computer Vision	Depth estimation, stereo vision, point cloud processing, and SLAM.	5
Real-World Applications	Case studies in autonomous vehicles, healthcare imaging, and facial recognition.	5
		End of table

Reference Books (Computer Vision)

Recommended Books - Computer Vision

No.	Book Title and Author
1	*Computer Vision: Algorithms and Applications* by Richard Szeliski
	Continued on next page

No.	Book Title and Author
2	*Deep Learning for Computer Vision* by Rajalingappaa Shanmugamani
3	*Pattern Recognition and Machine Learning* by Christopher M. Bishop
4	*Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
	End of table

List of Numerical Practicals (Computer Vision)

Practical Exercises - Computer Vision

No.	Practical Description
1	Implementing image filtering techniques: edge detection, blurring, and sharpening.
2	Feature extraction using HOG, SIFT, and ORB descriptors.
3	Training a CNN for image classification on CIFAR-10.
4	Implementing object detection with YOLO and Faster R-CNN.
5	Image segmentation using U-Net and Mask R-CNN.
6	Transfer learning with ResNet and EfficientNet.
7	Building a GAN for image synthesis.
8	Stereo vision and depth estimation using OpenCV.
9	Facial recognition using deep learning.
10	Case study: Applying computer vision in healthcare imaging.
	End of table

DS304E: Statistical Learning Theory (60 Lectures) (Back to Table)

Topic	Description	Hours
Introduction to Statistical Learning Theory	Basics of statistical learning, generalization, bias-variance trade-off, and empirical risk minimization.	5
Probably Approximately Correct (PAC) Learning	Definition of PAC learning, sample complexity, VC dimension, and learnability.	10
Vapnik-Chervonenkis (VC) Theory	VC dimension, uniform convergence, Rademacher complexity, and implications for model selection.	10
Kernel Methods in Machine Learning	Kernel trick, Support Vector Machines (SVMs), Reproducing Kernel Hilbert Space (RKHS), and Gaussian Processes.	10
Regularization and Capacity Control	Lasso, Ridge regression, elastic net, and controlling overfitting in learning models.	10
Concentration Inequalities	Hoeffding's inequality, Chernoff bounds, McDiarmid's inequality, and their role in statistical learning.	5
Statistical Foundations of Deep Learning	Neural network generalization, overparameterization, and information-theoretic perspectives on deep learning.	5
Advanced Topics and Applications	Online learning, adversarial learning, and applications in finance, healthcare, and NLP.	5

Reference Books

No.	Book Title and Author
1	*Understanding Machine Learning: From Theory to Algorithms* by Shai Shalev-Shwartz and Shai Ben-David
2	*The Nature of Statistical Learning Theory* by Vladimir Vapnik
3	*Foundations of Machine Learning* by Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar
4	*Mathematical Foundations of Machine Learning* by M. Anthony and P. L. Bartlett
5	*Kernel Methods for Pattern Analysis* by John Shawe-Taylor and Nello Cristianini

List of Numerical Practicals

No.	Practical Description
1	Implementing PAC learning algorithms and computing sample complexity.
2	Estimating the VC dimension of hypothesis classes in binary classification.
3	Applying the kernel trick in SVMs and kernel regression.
4	Computing Rademacher complexity and applying uniform convergence bounds.
5	Experimenting with regularization techniques in linear models and deep networks.
6	Implementing Gaussian Processes for probabilistic learning.
7	Using Hoeffding's inequality and Chernoff bounds for confidence intervals.
8	Analyzing generalization bounds in deep learning architectures.
9	Developing online learning algorithms with regret analysis.
10	Case study: Applying statistical learning theory to real-world datasets (e.g., healthcare, finance).

DS304F: Climate Data Science and Sustainable Resource Management Analytics (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Climate Data and Sources	Global and regional datasets: ERA5, CMIP6, CRU, IMD, MODIS; formats, APIs, resolution issues.	10
Climate Indicators	Temperature extremes, drought indices (SPI), rainfall variability, ENSO/monsoon indices.	10
Sustainable Resource Modeling	Water balance models, land use optimization, forest resource analytics, ecological footprints.	10
Climate Change Analytics	Detection and attribution, change-point analysis, trend slope estimation, long-term variability.	10
Carbon Accounting and Emissions Data	Emissions data sources, GHG inventories, carbon budget, lifecycle analysis.	10
Tools and Applications	Data analytics in SDG monitoring, decision support tools for sustainable farming and forestry.	10
		End of table

Recommended Books

No.	Book Title and Author
1	*Introduction to Modern Climate Change* by Andrew Dessler
2	*Climate Change and Climate Modeling* by J. David Neelin
3	*Sustainable Resource Management* by Walter Leal Filho and Vlasios Voudouris
4	*Environmental Modelling: Finding Simplicity in Complexity* by John Wainwright
5	*Data Analysis Methods in Physical Oceanography* by William J. Emery
	End of table

Numerical Practicals

No.	Practical Description
1	Importing and visualizing CMIP6 datasets (temperature, precipitation) using Python (xarray/netCDF).
2	Computing SPI (Standardized Precipitation Index) and rainfall anomalies using CRU or IMD data.
3	Estimating water demand and supply using simple water balance models in Excel or R.
4	Developing carbon emission estimates from fossil fuel consumption and transport data.
5	Trend analysis of long-term climate variables using Mann-Kendall Test and Sen's slope.
6	Optimization of land resource allocation using linear programming.
7	Energy-use modeling from IEA and FAOSTAT data sources using regression analysis.
8	Use of NDVI and vegetation indices for resource productivity assessment using MODIS data.
9	Lifecycle assessment (LCA) of products using openLCA or simplified spreadsheets.
10	Dashboard for monitoring SDG indicators using UN/World Bank datasets and visualization tools.
	End of table

DS305: Advanced Statistical Methods (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Multivariate Statistical Analysis	Mean vector and variance-covariance matrix, important properties, multivariate normal distribution, data reduction techniques, Hotelling's T^2, MANOVA.	10
Discriminant Analysis	Linear/Quadratic Discriminant Analysis (LDA, QDA), K-means, hierarchical clustering, spectral clustering.	10
Bayesian Statistics	Bayes' Theorem, prior/posterior distributions, Bayesian updating.	5
Markov Chain Monte Carlo (MCMC)	Gibbs Sampling, Metropolis-Hastings Algorithm, Bayesian hierarchical models.	5
Probabilistic Graphical Models	Bayesian Networks, applications in decision-making and predictive analytics.	5
Resampling Techniques	Bootstrap for confidence intervals, Jackknife resampling, hypothesis testing applications.	5
		Continued on next page...

Topic	Description	Hours
Monte Carlo Simulations	Importance sampling, variance reduction techniques, applications in finance and engineering.	5
Applications in Data Science	Case studies in healthcare analytics, financial modeling, customer retention, and reliability studies.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*The Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
2	*Bayesian Data Analysis* by Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin
3	*Monte Carlo Methods in Financial Engineering* by Paul Glasserman
	End of table

Practical Exercises

No.	Practical Description
1	Implementing Principal Component Analysis (PCA) for dimensionality reduction.
2	Applying Factor Analysis and Canonical Correlation Analysis.
3	Performing classification using Linear and Quadratic Discriminant Analysis (LDA, QDA).
4	Implementing clustering techniques (K-means, hierarchical clustering, spectral clustering).
5	Conducting Bayesian inference and updating priors using real-world datasets.
6	Running Gibbs Sampling and Metropolis-Hastings Algorithm for Bayesian analysis.
7	Conducting Bootstrap and Jackknife resampling techniques for confidence intervals.
8	Implementing Monte Carlo simulations for risk modeling and uncertainty quantification.
9	Building and interpreting Bayesian Networks for classification and prediction.
10	Case study: Advanced modeling using MCMC and multivariate techniques on real-world data.
	End of table

Semester IV

DS401: Big Data Analytics (60 Lectures) (Back to Table)

Topic	Description	Hours
Introduction to Big Data	Overview of big data, characteristics (Volume, Velocity, Variety, Veracity, and Value), and significance in modern analytics.	5
Hadoop Ecosystem	Understanding Hadoop architecture, HDFS, YARN, and MapReduce programming model.	5
Apache Spark Fundamentals	Spark architecture, Resilient Distributed Datasets (RDDs), DataFrames, and Spark SQL.	5
Distributed File Systems	HDFS, Amazon S3, and Google Cloud Storage for big data management.	5
		Continued on next page...

Topic	Description	Hours
Stream Processing Frameworks	Apache Kafka, Flink, and Spark Streaming for real-time analytics.	5
NoSQL Databases for Big Data	MongoDB, Cassandra, and HBase for scalable data storage and retrieval.	5
Statistical Methods in Big Data	Exploratory Data Analysis (EDA), descriptive statistics, and probability distributions in large datasets.	5
Hypothesis Testing in Big Data	T-tests, chi-square tests, ANOVA, and permutation testing for large-scale datasets.	5
Regression Models	Linear regression, logistic regression, ridge regression, and Lasso for large datasets.	5
Machine Learning for Big Data	Implementing ML models using Spark MLlib, feature engineering, and scalable learning algorithms.	5
Survival Analysis in Big Data	Kaplan-Meier estimators, Cox proportional hazards model, and applications in large-scale analytics.	5
Cloud Computing for Big Data	Cloud-based big data solutions using AWS, Google Cloud, and Azure.	5
Data Pipeline Architectures	ETL processes, batch processing vs. real-time processing, data lakes, and warehouses.	5
Performance Tuning	Optimizing big data applications: partitioning, caching, indexing, and tuning Spark jobs.	5
Industry Applications	Case studies on big data analytics in finance, healthcare, marketing, and cybersecurity.	5

No.	Book Title and Author
1	*Hadoop: The Definitive Guide* by Tom White
2	*Big Data: Principles and Best Practices of Scalable Real-Time Data Systems* by Nathan Marz
3	*Spark: The Definitive Guide* by Bill Chambers and Matei Zaharia
4	*Statistics for Big Data for Dummies* by Alan Anderson
5	*NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence* by Pramod J. Sadalage and Martin Fowler

List of Numerical Practicals

No.	Practical Description
1	Setting up and managing a Hadoop cluster, working with HDFS commands.
2	Writing and executing basic MapReduce programs in Java/Python.
3	Performing data transformations using Spark RDDs and DataFrames.
4	Implementing real-time data streaming using Apache Kafka and Spark Streaming.
5	Performing exploratory data analysis (EDA) on large datasets using Spark SQL.
6	Conducting hypothesis tests on big data using Spark and Python.
7	Training and evaluating regression models for large-scale datasets using Spark MLlib.
8	Working with NoSQL databases: CRUD operations in MongoDB and Cassandra.
9	Building scalable machine learning models and performing hyperparameter tuning.
10	Implementing survival analysis techniques on large datasets in Spark.

DS402: Discipline Elective III (Choose one) (Back to Table)

Students must choose one elective from the following options:

DS402A: Quality Control and Six Sigma (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Quality Control	Concepts of quality, quality assurance, and statistical process control (SPC).	5
Control Charts for Variables	X-bar, R, and S charts: construction, interpretation, and process capability analysis.	10
Control Charts for Attributes	p, np, c, and u charts with applications.	10
Process Capability Analysis	Cp, Cpk, Pp, Ppk indices and their interpretation.	5
Acceptance Sampling Plans	Single, double, and sequential sampling plans, OC curves, ASN functions.	10
Introduction to Six Sigma	DMAIC framework, roles and responsibilities, project selection.	5
Six Sigma Tools	Cause-and-effect diagram, Pareto analysis, FMEA, and control plans.	5
Lean Six Sigma	Lean principles, waste reduction, and integration with Six Sigma.	5
Case Studies and Implementation	Case studies from manufacturing and service industries.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Introduction to Statistical Quality Control* by Douglas C. Montgomery
2	*Statistical Quality Control* by Eugene Grant and Richard Leavenworth
3	*The Six Sigma Handbook* by Thomas Pyzdek and Paul Keller
4	*Lean Six Sigma for Service* by Michael L. George
	End of table

Practical Exercises

No.	Practical Description
1	Construct and interpret control charts for variables and attributes.
2	Conduct process capability analysis using real data.
3	Design acceptance sampling plans and plot OC curves.
4	Apply DMAIC on a selected quality improvement problem.
5	Perform FMEA for a process or product.
6	Use Pareto analysis to identify critical quality issues.
7	Implement Lean Six Sigma tools in a case study.
8	Use Minitab or R for quality control and Six Sigma analysis.
9	Simulate process control scenarios and interpret results.
10	Develop a control plan based on process data and customer needs.
	End of table

DS402B: Marketing and Social Media Analytics (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Marketing Analytics	Role in marketing decision-making, types of marketing data.	5
Customer Analytics	Segmentation, lifetime value, RFM analysis, churn modeling.	10
Digital and Social Media Metrics	Impressions, reach, CTR, conversions, engagement analysis.	10
Text Analytics for Social Media	Sentiment analysis, topic modeling, influencer detection.	10
Campaign Analytics	A/B testing, multivariate testing, ROI measurement.	10
Marketing Mix Modeling	Media attribution, marketing ROI, optimization of spend.	10
Dashboards and Visualization	Interactive reports using Tableau/Power BI or Python dashboards.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Marketing Analytics: Strategic Models and Metrics* by Stephan Sorger
2	*Social Media Analytics: Techniques and Insights* by Matthew Ganis and Avinash Kohirkar
3	*Digital Marketing Analytics* by Chuck Hemann and Ken Burbary
	End of table

Practical Exercises

No.	Practical Description
1	Analyze customer data for segmentation and LTV.
2	Perform churn prediction using classification models.
3	Collect and analyze real-time social media metrics.
4	Conduct sentiment analysis on Twitter/Facebook data.
5	Execute A/B testing simulations on marketing campaigns.
6	Build a marketing mix model with advertising spend data.
7	Design interactive dashboards for campaign reporting.
8	Case study: Evaluate marketing performance of a brand.
	End of table

DS402C: Statistical Genetics and Genomics (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Statistical Genetics	Basics of inheritance, Mendelian laws, Hardy-Weinberg equilibrium.	5
Genetic Association Studies	Case-control and cohort designs, linkage analysis, GWAS.	10
		Continued on next page...

Topic	Description	Hours
Quantitative Trait Loci (QTL) Analysis	Polygenic traits, marker-based analysis, heritability.	8
Gene Expression and Genomic Data	Microarray and RNA-Seq data analysis.	7
High-Dimensional Genomics	Multiple testing, penalized regression, dimension reduction.	7
Epigenetics and Gene Regulation	DNA methylation, histone modification, regulatory networks.	5
Single-cell Genomics	Technologies, data analysis, challenges and applications.	5
Pathway and Network Analysis	Gene set enrichment, biological networks.	5
Software and Tools	PLINK, Bioconductor, R packages for genetic data.	8
		End of table

Recommended Books

No.	Book Title and Author
1	*Statistical Genetics: Gene Mapping through Linkage and Association* by Benjamin Neale, Sarah Medland, Danielle Posthuma
2	*Genetics and Analysis of Quantitative Traits* by Lynch and Walsh
3	*Statistical Methods in Genetic Epidemiology* by Duncan Thomas
	End of table

Practical Exercises

No.	Practical Description
1	Test Hardy-Weinberg equilibrium for genotype data.
2	Perform case-control association study using PLINK.
3	Conduct linkage analysis and visualize results.
4	Analyze microarray or RNA-seq gene expression data.
5	Apply penalized regression to high-dimensional genomic data.
6	Conduct gene set enrichment analysis (GSEA).
7	Visualize genetic and epigenetic networks using R or Cytoscape.
8	Analyze single-cell RNA-seq data using Bioconductor packages.
9	Explore DNA methylation patterns using epigenetic datasets.
10	Integrative analysis: combine genomic, transcriptomic, and epigenomic data.
	End of table

DS402D: Reinforcement Learning (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Reinforcement Learning	Basics of RL, rewards, value functions, and exploration-exploitation trade-off.	5
		Continued on next page

Topic	**Description**	**Hours**
Markov Decision Processes (MDPs)	State transitions, policies, Bellman equations, and dynamic programming.	10
Model-Free RL	Monte Carlo methods, temporal difference learning, Q-learning, and SARSA.	10
Policy Gradient Methods	REINFORCE algorithm, actor-critic methods, and advantage estimation.	10
Deep Reinforcement Learning	Deep Q-Networks (DQN), experience replay, and target networks.	10
Multi-Agent Reinforcement Learning (MARL)	Cooperative and competitive environments, Nash equilibria, and self-play strategies.	5
Applications in Robotics	RL for robot control, motion planning, and autonomous navigation.	5
Game AI	RL in game environments, AlphaGo, AlphaZero, and applications in e-sports.	5
		End of table

Recommended Books

No.	**Book Title and Author**
1	*Reinforcement Learning: An Introduction* by Richard S. Sutton and Andrew G. Barto
2	*Deep Reinforcement Learning Hands-On* by Maxim Lapan
3	*Algorithms for Reinforcement Learning* by Csaba Szepesvári
4	*Multi-Agent Reinforcement Learning* by Lucian Busoniu, Robert Babuska, and Bart De Schutter
5	*Artificial Intelligence and Games* by Georgios N. Yannakakis and Julian Togelius
	End of table

List of Numerical Practicals

Practical Exercises

No.	**Practical Description**
1	Implementing value iteration and policy iteration for MDPs.
2	Developing a Q-learning agent for a grid-world environment.
3	Implementing Deep Q-Networks (DQN) for Atari game environments.
4	Training a reinforcement learning agent using policy gradient methods.
5	Developing an actor-critic model for continuous control tasks.
6	Implementing multi-agent reinforcement learning in a competitive game.
7	Applying RL to robotic motion planning using OpenAI Gym.
8	Building an RL model for autonomous trading in financial markets.
9	Implementing AlphaZero-style learning for board games.
10	Case study: Reinforcement learning for real-world decision-making problems.
	End of table

DS402E: Causal Inference and Counterfactual Analysis (60 Lectures) (Back to Table)

Topic	Description	Hours
Introduction to Causal Inference	Fundamentals of causality, correlation vs. causation, causal graphs, and applications in real-world problems.	5
Structural Causal Models (SCMs)	Directed acyclic graphs (DAGs), do-calculus, interventions, back-door and front-door criteria.	10
Potential Outcomes Framework	Rubin causal model, average treatment effect (ATE), conditional average treatment effect (CATE), and heterogeneous treatment effects.	10
Instrumental Variables (IV)	IV assumptions, two-stage least squares (2SLS), weak instruments, and applications in econometrics and epidemiology.	10
Propensity Score Methods	Matching, inverse probability weighting (IPW), overlap diagnostics, and covariate balance assessment.	10
Bayesian Causal Inference	Bayesian networks, Bayesian hierarchical models, and Bayesian approaches to causal discovery.	5
Advanced Topics in Causal Inference	Mediation analysis, synthetic control methods, difference-in-differences (DiD), and causal reinforcement learning.	5
Applications and Case Studies	Causal inference in healthcare, economics, social sciences, and machine learning.	5

No.	Book Title and Author
1	*Causality: Models, Reasoning, and Inference* by Judea Pearl
2	*The Book of Why: The New Science of Cause and Effect* by Judea Pearl and Dana Mackenzie
3	*Causal Inference: The Mixtape* by Scott Cunningham
4	*Counterfactuals and Causal Inference* by Stephen L. Morgan and Christopher Winship
5	*Bayesian Networks and Decision Graphs* by Finn V. Jensen and Thomas D. Nielsen

List of Numerical Practicals

No.	Practical Description
1	Implementing causal graphs and DAGs using the 'networkx' and 'causalgraphicalmodels' Python libraries.
2	Performing do-calculus and back-door adjustment for causal effect estimation.
3	Estimating treatment effects using the potential outcomes framework.
4	Applying instrumental variable regression using 2SLS in Python/R.
5	Conducting propensity score matching and inverse probability weighting (IPW).
6	Performing Bayesian causal inference using probabilistic graphical models.
7	Implementing synthetic control methods for causal impact assessment.
8	Using difference-in-differences (DiD) for policy evaluation.
9	Building causal reinforcement learning models for decision-making.
10	Case study: Applying causal inference methods to real-world datasets (e.g., healthcare, economics).

DS402F: Geospatial Analytics for Environmental Planning (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Basics of Geospatial Data	Types of spatial data (raster, vector), coordinate systems, spatial resolution, file formats, and metadata.	5
GIS and Remote Sensing Techniques	Image preprocessing, classification (supervised/unsupervised), change detection, NDVI, indices.	10
Urban and Regional Planning Models	LULC mapping, slope-aspect zoning, proximity/buffer analysis, transport network modeling.	10
Environmental Impact Mapping	Hazard risk zoning, noise/pollution mapping, ecological corridor design, protected area assessments.	10
Geostatistical Methods	Interpolation techniques (IDW, Kriging), variogram modeling, spatial autocorrelation, Moran's I.	10
Planning Case Studies	Watershed-based planning, smart city zoning, urban heat island mitigation using spatial data.	10
Spatial Decision Support Systems	GIS plugins, spatial querying, QGIS, GRASS, Python integration for automated spatial planning.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Remote Sensing and Image Interpretation* by Thomas Lillesand, Ralph Kiefer, Jonathan Chipman
2	*GIS Fundamentals: A First Text on Geographic Information Systems* by Paul Bolstad
3	*Geospatial Analysis: A Comprehensive Guide* by Michael de Smith, Michael Goodchild, Paul Longley
4	*Introductory Geographic Information Systems* by John R. Jensen
5	*Applied Geostatistics* by Edward H. Isaaks and R. Mohan Srivastava
	End of table

Numerical Practicals

No.	Practical Description
1	Importing and visualizing shapefiles and satellite images in QGIS.
2	NDVI computation using Landsat or Sentinel imagery and vegetation mapping.
3	Land use/land cover (LULC) classification using supervised learning in Google Earth Engine.
4	Spatial interpolation using Inverse Distance Weighting (IDW) and ordinary Kriging in R or QGIS.
5	Creating slope and aspect maps from a Digital Elevation Model (DEM).
6	Hotspot detection using Moran's I and Getis-Ord statistics.
7	Delineating watersheds using DEM and flow accumulation in QGIS/GRASS.
8	Buffer analysis for green belt and industrial zones around urban areas.
9	Land suitability analysis for urban expansion using weighted overlay.
10	Web map development using Leaflet or Python Folium for public-facing spatial planning tools.

End of table

DS403: Discipline Elective IV (Choose one) (Back to Table)

Students must choose one elective from the following options:

DS403A: Supply Chain Analytics (60 Lectures) (Back to Table)

Course Content

Topic	**Description**	**Hours**
Introduction to Supply Chains	Overview of supply chain components, objectives, and performance drivers.	5
Supply Chain Network Design	Location decisions, transportation models, distribution strategies.	10
Demand Forecasting	Quantitative techniques: time series, regression, machine learning-based methods.	10
Inventory Management	EOQ, safety stock, reorder point, and multi-echelon inventory optimization.	10
Supply Chain Coordination	Bullwhip effect, information sharing, contracts and incentives.	5
Logistics and Transportation Analytics	Vehicle routing, scheduling, and last-mile delivery optimization.	5
Supplier Selection and Risk Management	Evaluation techniques, risk modeling, supplier performance analytics.	5
Sustainability and Green Supply Chains	Environmental impact, reverse logistics, and carbon footprint reduction.	5
Analytics Tools	Use of R, Python, or specialized SCM software for modeling and simulation.	5
		End of table

Recommended Books

No.	**Book Title and Author**
1	*Designing and Managing the Supply Chain* by David Simchi-Levi, Philip Kaminsky, and Edith Simchi-Levi
2	*Supply Chain Management: Strategy, Planning, and Operation* by Sunil Chopra and Peter Meindl
3	*Logistics and Supply Chain Management* by Martin Christopher
4	*Business Analytics for Managers* by Gert H. N. Laursen and Jesper Thorlund
	End of table

Practical Exercises

No.	**Practical Description**
1	Design a basic supply chain network model using optimization.
2	Forecast demand using time series and regression techniques.
3	Simulate inventory management scenarios using EOQ and safety stock models.
4	Analyze the bullwhip effect and propose coordination strategies.
	Continued on next page

No.	Practical Description
5	Solve vehicle routing and scheduling problems.
6	Evaluate supplier performance using scoring and ranking techniques.
7	Build sustainability models to reduce supply chain environmental impact.
8	Use R or Python for simulation and visualization of supply chain flows.
9	Case study: Optimize a real-world supply chain using analytics.
10	Build an interactive dashboard for supply chain KPIs.
	End of table

DS403B: Financial Risk Analytics (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Financial Risk	Types of financial risk: market, credit, liquidity, operational.	5
Risk Measurement Techniques	Value at Risk (VaR), Expected Shortfall, Stress Testing.	10
Market Risk Models	Historical simulation, Monte Carlo, Delta-Normal methods.	10
Credit Risk Modeling	Credit scoring, probability of default, exposure at default, loss given default.	10
Portfolio Risk	Diversification, correlation, portfolio VaR.	5
Liquidity Risk Modeling	Measures of market and funding liquidity risk; liquidity-adjusted VaR.	5
Scenario and Sensitivity Analysis	What-if analysis, stress testing under extreme market conditions.	5
Regulatory Frameworks	Basel Accords, risk compliance, capital adequacy.	5
Risk Analytics Tools	R/Python for risk modeling, dashboarding, and real-time analytics.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Quantitative Risk Management* by McNeil, Frey, and Embrechts
2	*Financial Risk Manager Handbook* by Philippe Jorion
3	*Value at Risk: The New Benchmark for Managing Financial Risk* by Jorion
4	*The Essentials of Risk Management* by Michel Crouhy, Dan Galai, and Robert Mark
5	*Risk Management and Financial Institutions* by John C. Hull
	End of table

Practical Exercises

No.	Practical Description
1	Compute VaR using historical simulation, variance-covariance, and Monte Carlo methods.
2	Model credit risk using logistic regression and scorecard techniques.
3	Perform stress testing and scenario analysis for a sample portfolio.
	Continued on next page

No.	Practical Description
4	Simulate portfolio returns and compute portfolio-level VaR and CVaR.
5	Implement liquidity-adjusted risk measures and backtesting VaR.
6	Analyze diversification benefits through correlation matrices.
7	Evaluate capital requirements based on Basel III guidelines.
8	Develop interactive dashboards for real-time risk monitoring using R/Shiny or Python/Plotly.
9	Conduct sensitivity analysis to study the impact of market shocks.
10	Case study: Develop a risk analytics report for a financial institution.
	End of table

DS403C: Design of Clinical Trials (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Basics of Clinical Trials	Types of trials, phases, ethics and regulatory guidelines.	5
Trial Design and Randomization	Parallel, crossover, factorial designs; randomization techniques.	8
Sample Size Determination	Power analysis, Type I/II error, dropouts, interim analysis.	8
Endpoints and Outcomes	Primary/secondary endpoints, surrogate markers, composite endpoints.	8
Statistical Analysis Plans	Hypothesis testing, ITT vs PP, subgroup analysis.	8
Adaptive and Bayesian Designs	Sequential designs, adaptive randomization, Bayesian updates.	8
Monitoring and Reporting	DSMB, safety and efficacy, CONSORT guidelines.	5
Real-World Evidence and Pragmatic Trials	Use of electronic health records, pragmatic design principles.	5
Regulatory and Ethical Submissions	IND/IDE application, IRB/ethics committee review, trial registration.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Fundamentals of Clinical Trials* by Lawrence M. Friedman et al.
2	*Design and Analysis of Clinical Trials* by Shein-Chung Chow and Jen-Pei Liu
3	*Practical Biostatistics in Clinical Research* by Bertram K. C. Chan
	End of table

Practical Exercises

No.	Practical Description
1	Simulate randomization and blinding procedures.
2	Estimate sample size for different trial designs.
3	Analyze trial data using ITT and PP approaches.
	Continued on next page

No.	Practical Description
4	Perform interim analysis and interpret stopping rules.
5	Develop statistical analysis plans for mock trials.
6	Create trial monitoring dashboards.
7	Write a protocol summary for a clinical study.
8	Review CONSORT checklist for a published clinical trial.
9	Draft a submission plan for IRB/ethics approval.
10	Case study: Design a pragmatic trial using EHR data.
	End of table

DS403D: Cybersecurity Analytics (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Introduction to Cybersecurity Analytics	Overview of cybersecurity threats, attack vectors, and role of data analytics in security.	5
Intrusion Detection and Prevention Systems (IDPS)	Signature-based and anomaly-based detection using machine learning techniques.	10
Malware Classification Techniques	Static and dynamic malware analysis, behavioral modeling, and deep learning for classification.	10
Network Traffic Analysis	Packet inspection, anomaly detection, and feature extraction from network logs.	10
Threat Intelligence and Risk Assessment	Cyber threat intelligence, risk scoring models, and predictive analytics for security.	10
Digital Forensics and Incident Response	Log file analysis, evidence collection, and event correlation for cyber incident investigation.	5
Adversarial Machine Learning	Attack techniques on ML models and defense mechanisms against adversarial threats.	5
Real-World Applications	Case studies on intrusion detection, malware analysis, and fraud detection.	5
		End of table

Recommended Books

No.	Book Title and Author
1	*Cybersecurity Data Science: Applying Machine Learning and AI to Identify and Prevent Threats* by Scott Mongeau
2	*Machine Learning and Security* by Clarence Chio and David Freeman
3	*Practical Malware Analysis* by Michael Sikorski and Andrew Honig
4	*The Art of Memory Forensics* by Michael Hale Ligh, Andrew Case, Jamie Levy, Aaron Walters
	End of table

List of Numerical Practicals

Practical Exercises

No.	Practical Description
1	Implementing an Intrusion Detection System (IDS) using machine learning.
2	Classifying malware using static and dynamic analysis.
3	Analyzing network traffic logs to detect anomalies.
4	Performing feature engineering on cybersecurity datasets.
5	Applying deep learning techniques for malware classification.
6	Implementing risk assessment models for cybersecurity threats.
7	Conducting forensic analysis on log files for cyber incident detection.
8	Studying adversarial attacks on machine learning models and developing countermeasures.
9	Automating cybersecurity threat detection using SIEM tools.
10	Case study: Cyber fraud detection using real-world financial transaction data.
	End of table

DS403E: Advanced Statistical Computing (60 Lectures) (Back to Table)

Topic	Description	Hours
Monte Carlo Methods	Importance sampling, rejection sampling, and Monte Carlo integration techniques.	10
Markov Chain Monte Carlo (MCMC)	Gibbs Sampling, Metropolis-Hastings algorithm, and Hamiltonian Monte Carlo.	10
Stochastic Gradient Descent (SGD) Optimization	SGD, Adam, RMSprop, and optimization techniques for large-scale models.	10
GPU Programming for High-Performance Computing	CUDA programming, TensorFlow XLA, and PyTorch optimizations for GPU acceleration.	10
Distributed Computing	Parallel processing, cloud-based computing, and big data analytics frameworks (Hadoop, Spark).	5
Approximate Bayesian Inference	Variational inference, Expectation-Maximization (EM), and Bayesian optimization.	5
Best Practices in Statistical Software Development	Code optimization, reproducibility, and performance tuning in statistical computing.	5
Computational Efficiency in Statistical Methods	Efficient matrix operations, vectorization, and numerical stability techniques.	5
Real-World Applications	Case studies in finance, healthcare, and large-scale data analysis.	5

No.	Book Title and Author
1	*Bayesian Data Analysis* by Andrew Gelman
2	*Statistical Computing with R* by Maria L. Rizzo
3	*Monte Carlo Methods in Financial Engineering* by Paul Glasserman
4	*Pattern Recognition and Machine Learning* by Christopher M. Bishop

List of Numerical Practicals

No.	Practical Description
1	Implementing Monte Carlo simulations for probability estimation.
2	Applying Gibbs Sampling and Metropolis-Hastings algorithms.
3	Optimizing statistical models using stochastic gradient descent (SGD).
4	Running distributed computing tasks with Apache Spark and Hadoop.
5	Implementing Bayesian optimization for hyperparameter tuning.
6	Programming GPU-accelerated statistical models using CUDA and PyTorch.
7	Parallelizing matrix operations and statistical computing workflows.
8	Developing efficient Markov Chain Monte Carlo (MCMC) models.
9	Conducting large-scale financial risk analysis using Monte Carlo methods.
10	Case study: High-performance statistical computing in genomics and bioinformatics.

DS403F: Climate Risk Modeling and Impact Assessment (60 Lectures) (Back to Table)

Course Content

Topic	Description	Hours
Climate Hazard Modeling	Modeling extreme events (flood, heatwave, cyclone) using historical and reanalysis datasets; probability and return periods.	10
Vulnerability and Exposure Assessment	Measuring exposure to climate threats, socio-economic vulnerability indices, sensitivity and adaptive capacity.	10
Integrated Risk Index Development	Framework for risk index construction using PCA, normalization, and weighted summation.	10
Impact Modeling in Key Sectors	Modeling climate impacts in agriculture, water resources, health, and infrastructure.	10
Scenario Analysis and Projections	Use of RCPs, SSPs, and CMIP6 climate models for impact forecasting and uncertainty quantification.	10
Climate Finance and Policy Integration	Economic damage modeling, adaptation costing, risk transfer mechanisms, links to NDMA, UNFCCC, and IPCC frameworks.	10
		End of table

Recommended Books

No.	Book Title and Author
1	*Managing Climate Risk in the U.S. Financial System* – U.S. CFTC Climate-Related Market Risk Subcommittee Report
2	*Risk Modeling for Hazards and Disasters* by G. Woo
3	*Climate Risk: Economic and Financial Impact* by Mark Carey
4	*Climate Change and Disaster Risk Management* by Walter Leal Filho
5	*Climate Change 2022: Impacts, Adaptation and Vulnerability* – IPCC Working Group II Report
	End of table

Numerical Practicals

No.	Practical Description
1	Estimating return periods for extreme rainfall events using Gumbel distribution.
2	Generating flood hazard maps using satellite rainfall and DEM data in QGIS.
3	Developing heatwave risk indices using ERA5 temperature data and population density.
4	Creating a multi-dimensional vulnerability index using PCA from socio-economic indicators.
5	Modeling agricultural yield losses due to temperature/rainfall extremes using regression.
6	Performing uncertainty analysis using ensemble CMIP6 projections.
7	Projecting future flood/drought frequency using downscaled RCP scenarios.
8	Economic damage estimation using sectoral exposure data (e.g., transport, housing).
9	Climate risk dashboard development with real-time data (e.g., IMD, NOAA, World Bank).
10	Case study: Vulnerability assessment for a coastal district integrating hazards, exposure, and adaptation.
	End of table

DS404: Advanced Deep Learning and AI Innovations (60 Lectures) (Back to Table)

Topic	Description	Hours
Advanced Deep Learning Architectures	Exploration of transformer models (GPT, BERT, T5, Vision Transformers), GANs, VAEs, self-supervised learning (SimCLR, MoCo), and federated learning.	10
Advanced Reinforcement Learning	Deep Deterministic Policy Gradient (DDPG), Proximal Policy Optimization (PPO), Monte Carlo Tree Search (MCTS), and applications in robotics, trading, and autonomous systems.	10
Explainable AI and Trustworthy ML	AI ethics, bias detection, fairness-aware ML, adversarial robustness, and causal inference techniques.	10
Cutting-Edge Topics in AI Research	Large Language Models (LLMs), quantum machine learning, AI in drug discovery and healthcare, neuromorphic computing, and brain-inspired AI.	10
AI Productization and Scalability	Scalable AI architectures, MLOps (Apache Airflow, Kubeflow), Edge AI, TinyML, and cloud-based deployment strategies (Kubernetes, cloud infrastructure).	10
Real-World Applications	Case studies on AI in finance, healthcare, autonomous systems, and generative AI.	10

No.	Book Title and Author
1	*Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
2	*Grokking Deep Learning* by Andrew W. Trask
3	*Advances in Financial Machine Learning* by Marcos López de Prado
4	*Reinforcement Learning: An Introduction* by Richard S. Sutton and Andrew G. Barto
5	*Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* by Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, Klaus-Robert Müller

List of Numerical Practicals

No.	Practical Description
1	Implementing transformer-based NLP models (BERT, GPT) for text generation and classification.
2	Training and evaluating Generative Adversarial Networks (GANs) for image synthesis.
3	Applying Variational Autoencoders (VAEs) for unsupervised data generation.
4	Implementing reinforcement learning algorithms (DDPG, PPO) for autonomous decision-making.
5	Analyzing model fairness and detecting biases in AI models using fairness-aware ML techniques.
6	Exploring adversarial robustness by testing AI models against adversarial attacks.
7	Developing a quantum machine learning model using quantum computing frameworks.
8	Deploying deep learning models at scale using MLOps tools (Apache Airflow, Kubeflow).
9	Implementing Edge AI models for IoT and low-power device applications.
10	Case study: AI-powered financial fraud detection using deep learning.

DS405: Project (Back to Table)

This course focuses on research methodologies and project execution. Topics include formulating a research problem, conducting literature reviews, experimental design, hypothesis formulation, and data collection strategies. Students will also learn about model development and evaluation, research ethics, plagiarism policies, scientific paper writing, and effective visualization of research findings. The course provides best practices for thesis writing and guidelines for final project submission and evaluation.

Compulsory Internship (External/Internal) after Semester II

(Back to Table)

Project Evaluation Criteria (Final Semester Project – DS405)

Project Evaluation Scheme (Total: 100%)

S.No.	Evaluation Component	Weightage (%)
1	Problem Statement and Objectives: Clarity, relevance, and scope of the problem; alignment with data science domain.	10%
2	Literature Review and Background: Understanding of prior work, appropriate referencing, and contextual grounding.	10%
3	Data Acquisition and Exploration: Relevance of dataset, data understanding, quality checks, and EDA techniques.	10%
4	Methodology / Model Building: Selection and implementation of suitable models/techniques (ML, stats, etc.), justification, and alternatives.	25%
5	Results and Interpretation: Analytical insights, validation metrics, error analysis, and interpretation of outputs.	15%
6	Innovation and Originality: Novelty in approach, creative use of tools/methods, and intellectual contribution.	10%
7	Report and Documentation: Structure, clarity, presentation, reproducibility, visuals, and bibliography.	10%

(Continued on next page)

(Continued from previous page)

S.No.	Evaluation Component	Weightage (%)
8	Viva-Voce / Presentation: Oral communication, technical clarity, and defense of methodology and results.	10%
	Total	**100%**

Expected Outcomes

1. **Practical Experience**: Students will gain industry/academic experience through internships.
2. **Research Skills**: Students will learn scientific methods, experimentation, and technical writing.
3. **Innovative Project Development**: Application of data science methodologies to real-world challenges.
4. **Thesis & Paper Publication**: High-quality research output that can be published in conferences/journals.

Reference Books:

1. *Research Methods for Data Science* by Uwe Flick
2. *The Craft of Research* by Wayne C. Booth