Hindi Seva Mandal's
Shri Sant Gadge Baba
College of Engineering & Technology,
Near Z.T.C., Bhusawal. Dist. - Jalgaon (Maharashtra) Pin - 425203.

An Autonomous Institute Accreditation by NAAC with A+ Grade

(Affiliated to Dr. Babasaheb Ambedkar Technological University, Lonere.)

www.ssgbcoet.com

Bucket of Double Minor Specialization (DMS) Courses

Offers By

Bachelor of Technology (B. Tech.)

in

Computer Science and Engineering

With Effect from the Academic Year 2025-2026

4-year, 8 Semester Full time Programme Choice Based Credit

System (CBCS) and Grading System Outcome Based Education Pattern
Aligned with

National Education Policy (NEP) 2020

ADDITIONAL CRITERIA FOR DOUBLE MINOR (SPECIALIZATION) (With Effective from 2026-2027)										
Category under NEP	**Course Code**	**Course Title**	**Teaching Scheme**			**Evaluation Scheme**				**No. of Credits**
			L	**T**	**P**	**CA**	**MSE**	**ESE**	**Total**	
DMS	BTCS25DMS333	Introduction to Algorithms and Analysis	3	0	0	20	20	60	100	3
DMS	BTCS25DMS444	Probability Theory for Data Science	3	0	0	20	20	60	100	3
DMS	BTCS25DMS555	Data Analytics with Python	3	0	0	20	20	60	100	3
DMS	BTCS25DMS666	Introduction to Large Language Models	3	0	0	20	20	60	100	3
DMS	BTCS25DMS777	Machine Learning for Engineering and Science Applications	3	0	0	20	20	60	100	3
DMS	BTCS25DMS888	Statistical foundation of Big Data Analysis	3	0	0	20	20	60	100	3

Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme				No. of Credits
			L	T	P	CA	MSE	ESE	Total	
DMS	BTCS25DMS333	Introduction to Algorithms and Analysis	3	-	-	20	20	60	100	3

Syllabus BTCS25DMS333 Introduction to Algorithms and Analysis

Course Description:

This course introduces fundamental concepts of algorithm design and analysis, including asymptotic notations and recurrence relations. It covers key design techniques such as divide-and-conquer and dynamic programming for efficient problem solving. The course explores sorting, searching, and advanced data structures like trees, hashing, and heaps.It also includes graph algorithms such as traversal, minimum spanning trees, and shortest path problems. Advanced topics like amortized analysis, network flow, and computational complexity are discussed to understand efficiency and limitations of algorithms.

Course Objectives:

1. To understand fundamental concepts of algorithm design, analysis, and asymptotic notations for evaluating performance.
2. To apply algorithm design techniques such as divide-and-conquer, greedy methods, and dynamic programming to solve computational problems.
3. To study and implement efficient algorithms and data structures including sorting, searching, trees, hashing, and graph algorithms.
4. To analyze advanced concepts such as amortized analysis, network flow, and computational complexity to evaluate algorithm efficiency and limitations.

Course Outcomes:

CO1: Analyze and evaluate the performance of algorithms using asymptotic notations and recurrence relations.
CO2: Apply appropriate algorithm design techniques such as divide-and-conquer and dynamic programming to solve computational problems.
CO3: Implement and analyze efficient data structures and algorithms including sorting, searching, trees, hashing, and graph algorithms.
CO4: Demonstrate understanding of advanced concepts such as amortized analysis, network flow, and computational complexity in solving real-world problems.

CO–PO Mapping Matrix

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	3	2	2	1	-	-	-	-	-	-	2
CO2	3	3	3	2	2	-	-	-	1	-	-	2
CO3	3	3	3	3	2	-	-	-	1	-	-	2
CO4	3	3	2	2	1	-	-	-	-	-	-	3

Unit I: Algorithm Analysis and Divide & Conquer **(09 Hours)**
Sorting problem, time complexity (best case, average case, worst case), asymptotic analysis (Big-O, Omega, Theta), solving recurrences (substitution method, iteration method, recursion tree), divide-and-conquer strategy, applications of divide-and-conquer

Unit II: Sorting Techniques and Order Statistics **(09 Hours)**
Quick sort (algorithm and analysis), heap sort (max heap, min heap, heap operations), decision tree model for sorting, linear time sorting (counting sort, radix sort, bucket sort), order statistics (selection problem, median finding)

Unit III: Trees and Advanced Data Structures **(09 Hours)**
Hash functions, collision resolution techniques, binary search tree (BST), BST sort, randomly built BST, red-black tree, augmentation of data structures, van Emde Boas tree

Unit IV: Graph Algorithms and Dynamic Programming **(09 Hours)**

Dynamic programming (principles, optimal substructure, overlapping subproblems), graph representation, minimum spanning tree (Prim's algorithm), breadth-first search (BFS), depth-first search (DFS), shortest path algorithms (Dijkstra's algorithm, Bellman-Ford algorithm)

Unit V: Advanced Graphs, Amortized Analysis and Complexity **(09 Hours)**

All pairs shortest path problem, Floyd-Warshall algorithm, Johnson's algorithm, amortized analysis (aggregate, accounting, potential methods), disjoint set data structure (union-find), network flow, computational complexity (P, NP, NP-complete)

Text Books:

1. Introduction to Algorithms, Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, 3rd Edition, MIT Press.
2. Algorithm Design, Jon Kleinberg, Éva Tardos, Pearson Education.
3. Data Structures and Algorithm Analysis in C++, Mark Allen Weiss, Pearson.
4. Fundamentals of Computer Algorithms, Ellis Horowitz, Sartaj Sahni, Universities Press.

Reference Books

1. The Algorithm Design Manual, Steven S. Skiena, Springer.
2. Algorithms, Robert Sedgewick, Kevin Wayne, Addison-Wesley.
3. Introduction to the Design and Analysis of Algorithms, Anany Levitin, Pearson.
4. Data Structures and Algorithms Made Easy, Narasimha Karumanchi, CareerMonk Publications.
5. Algorithmics: The Spirit of Computing, David Harel, Yishai Feldman, Addison-Wesley.

NPTEL Course: **Introduction to Algorithms and Analysis, IIT Kharagpur, Prof. Sourav Mukhopadhyay.**
Link- https://nptel.ac.in/courses/106105164

Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme				No. of Credits
			L	T	P	CA	MSE	ESE	Total	
DMS	BTCS25DMS444	Probability Theory for Data Science	3	-	-	20	20	60	100	3

Syllabus - BTCS25DMS444 Probability Theory for Data Science

Course Description:

This course introduces the fundamental concepts of probability and statistical analysis. It covers random variables, probability distributions, and statistical measures essential for data interpretation. Students will learn estimation techniques and hypothesis testing methods. The course also explores correlation and regression for real-world data analysis. It builds a strong foundation for advanced data science and analytics applications.

Course Objectives:

1. To understand the fundamental concepts of probability and random variables.
2. To analyze probability distributions and statistical measures.
3. To apply estimation and hypothesis testing techniques in practical problems.
4. To interpret relationships between variables using correlation and regression.

Course Outcomes:

CO1: Explain probability concepts and random variables.
CO2: Analyze statistical distributions and compute moments.
CO3: Apply estimation and hypothesis testing techniques.
CO4: Evaluate relationships using correlation and regression analysis.

CO–PO Mapping Matrix

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	2	2	3	1	-	-	1	-	-	2
CO2	3	3	2	3	3	-	-	-	1	-	-	2
CO3	2	3	3	2	3	-	-	-	2	1	-	2
CO4	3	3	2	3	3	-	-	-	1	-	-	2
CO5	2	2	2	2	3	1	1	1	2	2	2	2

Unit I: Fundamentals of Probability **(09 Hours)**
Basics of Probability, Sample Space, Events, Axioms of Probability, Conditional Probability, Bayes' Theorem, Random Variables (Discrete and Continuous), Probability Distribution Functions.

Unit II: Random Variables and Moments **(09 Hours)**
Mathematical Expectation, Moments (Raw and Central), Variance, Standard Deviation, Moment Generating Functions, Chebyshev's Inequality, Markov's Inequality

Unit III: Probability Distributions **(09 Hours)**
Standard Distributions: Binomial, Poisson, Normal Distributions, Properties and Applications, Higher Dimensional Distributions, Joint, Marginal and Conditional Distributions.

Unit IV: Functions and Sampling Theory **(09 Hours)**
Functions of Several Random Variables, Transformation of Variables, Cross Moments, Covariance, Correlation, Limiting Distributions (Law of Large Numbers, Central Limit Theorem), Descriptive Statistics, Sampling Distributions.

Unit V: Statistical Inference and Regression **(09 Hours)**
Point Estimation, Interval Estimation, Testing of Hypothesis (Z-test, t-test, Chi-square test), Errors in Testing, Analysis of Correlation, Linear Regression and its Applications.

Text Books:

1. Sheldon M. Ross – Introduction to Probability and Statistics for Engineers and Scientists
2. S. C. Gupta and V. K. Kapoor – Fundamentals of Mathematical Statistics

3. William Mendenhall & Robert J. Beaver – Introduction to Probability and Statistics

Reference Books

1. Jay L. Devore – *Probability and Statistics for Engineering and the Sciences*
2. Ronald E. Walpole et al. – *Probability and Statistics for Engineers and Scientists*
3. Hogg, McKean & Craig – *Introduction to Mathematical Statistics*

NPTEL Course: **Introduction to Probability Theory and Statistics, IIT Delhi** Prof. S Dharmaraja
Link - https://nptel.ac.in/courses/111102160

Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme				No. of Credits
			L	T	P	CA	MSE	ESE	Total	
DMS	BTCS25DMS555	Data Analytics with Python	3	-	-	20	20	60	100	3

Syllabus - BTCS25DMS555 Data Analytics with Python

Course Description:

This course introduces fundamental concepts of data analytics using Python. It covers probability, sampling theory, hypothesis testing, regression models, model evaluation techniques, and clustering methods. The course emphasizes statistical reasoning, predictive modeling, and hands-on analytical skills required for data-driven decision-making.

Course Objectives:

1. To introduce data analytics concepts and Python fundamentals for data analysis.
2. To develop understanding of probability, sampling, and hypothesis testing techniques.
3. To build skills in regression modeling and statistical inference.
4. To introduce classification, clustering, and model evaluation techniques.
5. To enable students to apply analytical methods for solving real-world problems.

Course Outcomes:

CO1: Apply Python tools for data preprocessing and exploratory analysis.
CO2: Use probability, sampling, and hypothesis testing methods for statistical analysis.
CO3: Develop and evaluate regression models for predictive analytics.
CO4: Apply classification and clustering techniques to analyze datasets.
CO5: Interpret analytical results and build data-driven models for decision-making.

CO–PO Mapping Matrix

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	2	1	3	2	2	1	–	1	–	1	–	2
CO2	3	3	–	2	2	–	–	–	1	–	–	2
CO3	2	2	2	3	3	–	1	–	–	–	–	2
CO4	2	2	2	3	3	–	1	–	–	–	–	2
CO5	1	1	2	3	3	1	1	2	1	1	1	3

Unit I: Introduction to Data Analytics and Probability **(09 Hours)**
Introduction to Data Analytics, Data Science Lifecycle, Python Fundamentals for Data Analytics (NumPy, Pandas basics), Introduction to Probability, Basic Probability Rules and Random Variables

Unit II: Sampling and Hypothesis Testing **(09 Hours)**
Sampling Methods, Sampling Distributions, Central Limit Theorem, Hypothesis Testing Framework, p-value and Confidence Intervals

Unit III: ANOVA and Regression Analysis **(09 Hours)**
Two-Sample Testing, One-Way ANOVA, Two-Way ANOVA, Simple Linear Regression, Multiple Linear Regression, Model Assumptions and Evaluation

Unit IV: Advanced Regression and Model Evaluation **(09 Hours)**
Maximum Likelihood Estimation (MLE), Logistic Regression, ROC Curve and Model Performance Metrics, Regression Model Building Process

Unit V: Statistical Tests and Machine Learning Techniques **(09 Hours)**
Chi-Square Test, Introduction to Cluster Analysis, Clustering Techniques (K-Means Concept), Classification and Regression Trees (CART), Applications of Clustering and Classification

Textbooks

1. *Practical Statistics for Data Scientists* – Peter Bruce & Andrew Bruce
2. *An Introduction to Statistical Learning* – Gareth James et al.
3. *Python for Data Analysis* – Wes McKinney

Reference Books

1. *Applied Multivariate Statistical Analysis* **–** Richard A. Johnson & Dean W. Wichern
2. *The Elements of Statistical Learning* **–** Hastie, Tibshirani & Friedman
3. *Think Stats* **–** Allen B. Downey

NPTEL Course: Data Analytics with Python, IIT Roorkee, Prof. A Ramesh
Link- https://nptel.ac.in/courses/106107220
Link- https://nptel.ac.in/courses/110106072

Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme				No. of Credits
			L	T	P	CA	MSE	ESE	Total	
DMS	BTCS25DM666	Introduction To Large Language Models (LLMS)	3	-	-	20	20	60	100	3

Syllabus - BTCS25DM666 Introduction To Large Language Models (LLMS)

Course Description:

This course provides an in-depth understanding of Natural Language Processing (NLP) with a focus on modern deep learning and generative AI techniques. It covers statistical and neural language models, transformers, and large language models. Students will learn advanced concepts such as attention mechanisms, prompt engineering, and retrieval-augmented generation. The course emphasizes practical implementation using frameworks like PyTorch and HuggingFace for real-world NLP applications.

Course Objectives:

1. To introduce fundamental concepts and techniques of Natural Language Processing.
2. To understand statistical and neural language models for text processing.
3. To develop knowledge of deep learning architectures such as RNNs, CNNs, and Transformers.
4. To explore modern NLP techniques including large language models, prompting, and retrieval methods.
5. To enable practical implementation of NLP applications using contemporary tools and frameworks.

Course Outcomes:

CO1 Explain fundamental concepts, NLP pipeline, and language modeling techniques.
CO2: Apply statistical and neural models for text representation and processing.
CO3: Analyze deep learning architectures such as RNNs, CNNs, and Transformers in NLP tasks.
CO4: Evaluate advanced NLP techniques including LLMs, prompting, and retrieval-based systems.
CO5: Design and implement NLP applications using modern frameworks and tools.

CO–PO Mapping Matrix

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	-	-	1	-	-	-	-	-	-	2
CO2	3	3	2	2	2	-	-	-	-	-	-	2
CO3	3	3	2	3	2	-	-	-	-	-	-	2
CO4	3	3	3	3	2	-	-	-	-	-	-	2
CO5	3	3	3	3	3	-	-	-	2	2	1	3

Unit I: Introduction to NLP and Language Modeling **(09 Hours)**
Course Introduction, NLP Pipeline, Applications of NLP, Introduction to Statistical Language Models, N-gram Models, Smoothing Techniques, Evaluation of Language Models

Unit II: Deep Learning and Word Representation **(09 Hours)**
Perceptron and ANN, Backpropagation, Convolutional Neural Networks (CNN), Introduction to PyTorch, Word2Vec and fastText, GloVe, Tokenization Strategies

Unit III: Neural Language Models and Sequence Learning **(09 Hours)**
Neural Language Models (CNN, RNN), LSTM and GRU, Sequence-to-Sequence Models, Greedy Decoding and Beam Search, Sampling Techniques (Top-k, Temperature, Nucleus), Attention Mechanism, Applications of Sequence Models

Unit IV: Transformers and Large Language Models **(09 Hours)**
Introduction to Transformers, Self Attention and Multi-Head Attention, Positional Encoding and Layer Normalization, Transformer Implementation using PyTorch, Pre-training: ELMo and BERT, Encoder-Decoder & Decoder-only Models, HuggingFace and Introduction to Prompting.

Unit V: Advanced NLP, Retrieval and Ethics **(09 Hours)**

Instruction Tuning and Prompt Engineering, Reinforcement Learning from Human Feedback (RLHF), Retrieval-Augmented Generation (RAG), Knowledge Graphs and KGQA, Parameter-efficient Fine-tuning (LoRA, Prefix, Prompt Tuning), Modern LLMs (GPT, LLaMA, Gemini, etc.), Ethical NLP (Bias, Fairness, Toxicity)

Textbooks

1. Speech and Language Processing – Daniel Jurafsky and James H. Martin Comprehensive coverage of NLP fundamentals, language models, and applications.
2. Natural Language Processing with Python – Steven Bird, Ewan Klein, Edward Loper Practical introduction using Python and NLTK.
3. Deep Learning – Ian Goodfellow, Yoshua Bengio, Aaron Courville Covers neural networks, CNNs, RNNs, and deep learning foundations.

Reference Books

1. Natural Language Processing with Transformers – Lewis Tunstall, Leandro von Werra, Thomas Wolf Focus on transformers, BERT, and HuggingFace implementation.
2. Transformers for Natural Language Processing – Denis Rothman Detailed explanation of transformer models and applications.
3. Speech and Language Processing (3rd Edition Draft) – Jurafsky & Martin Updated content on transformers and modern NLP techniques.

NPTEL Course: Introduction to Large Language Models (LLMs), IIT Delhi IIT Bombay, Prof. Tanmoy Chakraborty, Prof. Soumen Chakraborti

Link - https://nptel.ac.in/courses/106102576

Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme				No. of Credits
			L	T	P	CA	MSE	ESE	Total	
PCC	BTCS25DMS777	Machine Learning For Engineering And Science Applications	3	-	-	20	20	60	100	3

Syllabus - BTCS25DMS777 Machine Learning For Engineering And Science Applications

Course Description:

This course introduces the fundamental concepts and techniques of Machine Learning, including mathematical foundations such as linear algebra, probability, and optimization. It covers supervised learning methods like regression, along with model evaluation concepts such as bias-variance tradeoff and regularization. The course further explores neural networks and deep learning architectures including CNNs and RNNs for complex data analysis. It also includes classical and advanced machine learning techniques, emphasizing practical implementation for real-world applications.

Course Objectives:

1. To introduce fundamental concepts and mathematical foundations of Machine Learning.
2. To develop understanding of supervised learning techniques such as regression and classificati
3. To provide knowledge of neural networks and deep learning architectures.
4. To familiarize students with classical and advanced machine learning algorithms.
5. To enable practical implementation and evaluation of machine learning models for real-world problems.

Course Outcomes:

CO1: Explain fundamental concepts, mathematical foundations, and applications of machine learning
CO2: Apply regression and classification techniques to solve real-world problems.
CO3: Analyze neural networks and deep learning models such as CNNs and RNNs.
CO4: Evaluate performance of machine learning models using appropriate metrics and techniques .
CO5: Design and implement machine learning solutions using modern tools and algorithms.

CO–PO Mapping Matrix

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	-	-	1	-	-	-	-	-	-	2
CO2	3	3	2	2	2	-	-	-	-	-	-	2
CO3	3	3	2	2	2	-	-	-	-	-	-	2
CO4	3	3	3	3	2	-	-	-	-	-	-	2
CO5	3	3	3	3	3	-	-	-	2	2	1	3

Unit I: Mathematical and Computational Foundations **(09 Hours)**
Introduction to Machine Learning: Definition, types, and applications, Linear Algebra: Vectors, matrices, eigenvalues, eigenvectors, Probability Theory: Random variables, distributions, expectation, variance, Numerical Computation and Optimization, Introduction to Machine Learning Libraries/Packages (e.g., Python-based tools).

Unit II: Regression Techniques and Model Evaluation **(09 Hours)**
Linear Regression: Model formulation and assumptions, Logistic Regression: Classification and sigmoid function, Bias-Variance Tradeoff, Regularization Techniques (L1, L2), Gradient Descent Variants (Batch, Stochastic, Mini-batch), Maximum Likelihood Estimation (MLE) and Maximum A Posteriori (MAP), Applications of Regression Models

Unit III: Neural Networks and Deep Learning Fundamentals **(09 Hours)**
Introduction to Neural Networks, Multilayer Perceptron (MLP), Backpropagation Algorithm, Activation Functions, Applications of Neural Networks

Unit IV: Deep Learning Architectures **(09 Hours)**
Convolutional Neural Networks (CNN): CNN Operations (Convolution, Pooling), CNN Architectures, Training and Optimization, Transfer Learning, Recurrent Neural Networks (RNN): RNN, LSTM, GRU, Sequence Modeling, Applications of CNNs and RNNs

Unit V: Classical and Advanced Machine Learning Techniques **(09 Hours)**
Classical Machine Learning Algorithms: Bayesian Regression, Decision Trees and Random Forests, Support Vector Machines (SVM), Naïve Bayes, Clustering Techniques: k-Means, k-Nearest Neighbors (kNN), Gaussian Mixture Models (GMM), Expectation Maximization (EM), Advanced Techniques: Structured Probabilistic Models, Monte Carlo Methods, Autoencoders, Generative Adversarial Networks (GANs), Applications of Classical and Advanced Techniques.

Reference Books

1. Machine Learning – Tom M. Mitchell Classic introductory textbook covering fundamentals of ML algorithms
2. Pattern Recognition and Machine Learning – Christopher M. Bishop Strong mathematical foundation; widely used for probability-based ML.
3. Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow – Aurélien Géron Practical implementation with Python and real-world examples.

NPTEL Course: Machine Learning for Engineering and science applications, **By Prof. Balaji Srinivasan, Prof. Ganapathy Krishnamurthi | IIT Madras**
Link- https://onlinecourses.nptel.ac.in/noc26_cs76/preview

Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme				No. of Credits
			L	T	P	CA	MSE	ESE	Total	
DMS	BTCS25DMS888	Statistical foundation of Big Data Analysis	3	-	-	20	20	60	100	3

Syllabus - BTCS25DMS888 Statistical foundation of Big Data Analysis

Course Description:

This course introduces the fundamental concepts of Big Data and statistical learning techniques used for analyzing large and complex datasets. It covers statistical inference methods such as estimation and hypothesis testing. The course emphasizes model evaluation through bias-variance analysis. It also explores multivariate analysis and regression techniques. Advanced topics include clustering, PCA, and high-dimensional data analysis.

Course Objectives:

1. Develop a clear understanding of Big Data concepts, challenges, and analytical frameworks.
2. Build strong foundations in statistical inference methods such as estimation and hypothesis testing.
3. Introduce principles of statistical learning theory, including bias-variance tradeoff.
4. Equip students with knowledge of multivariate analysis and regression models.
5. Enable understanding of high-dimensional data analysis techniques like clustering and PCA.

Course Outcomes:

CO1: Explain Big Data concepts and distinguish between data processing and data analytics approaches.
CO2: Apply statistical inference techniques (MLE, method of moments, hypothesis testing) to real-world datasets.
CO3: Analyze model performance using bias, variance, and mean squared error concepts.
CO4: Implement multivariate regression models and interpret their results in practical scenarios.
CO5: Apply unsupervised learning methods such as clustering and Principal Component Analysis (PCA) for high-dimensional data.

CO–PO Mapping Matrix

COs \ POs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	–	–	1	–	–	–	–	1	–	1
CO2	3	3	–	2	2	–	–	–	–	1	–	1
CO3	3	3	1	2	2	–	–	–	–	1	–	1
CO4	3	3	2	2	3	–	–	–	–	1	1	1
CO5	3	3	2	2	3	–	–	–	–	1	1	1

Unit 1: Introduction to Big Data and Statistical Foundations (09 Hours)

What is Big Data? Examples and case studies of Big Data, Challenges in handling Big Data Big Data Analysis vs Big Data Processing, Role of high-dimensional statistics in Big Data Overview of statistical framework, Classical statistical analysis (linear algebra & calculus-based inference) ,Big Data analysis using linear algebra, calculus, and high-performance computing.

Unit 2: Fundamentals of Statistical Inference (09 Hours)

Unbiased Estimation, Maximum Likelihood Estimation (MLE) ,Method of Moments Estimation Asymptotic properties and case studies , Concept of interval estimation with examples Basics of hypothesis testing, Real-life examples and case studies

Unit 3: Statistical Learning Theory **(09 Hours)** Bias, Variance, and Mean Squared Error (MSE) , Real-life examples and case studies Bias-Variance Tradeoff , Comparison of estimation models based on bias and variance Case studies on model performance
Unit 4: Multivariate Analysis and Modeling **(09 Hours)** Regression and Multivariate Regression, Gauss-Markov Theorem and applications, Introduction to multivariate distributions , Multivariate Normal Distribution (PDF, CDF) ,Conditional distributions Linear and affine transformations ,Other common multivariate distributions.
Unit 5: Advanced Topics in Big Data Analytics **(09 Hours)** Unsupervised learning concepts, Hierarchical and non-hierarchical clustering ,K-means clustering (detailed study) ,Curse of dimensionality ,Volume asymptotics and impact on clustering Population PCA: concepts and applications ,Sample PCA: concepts and applications .Network data concepts, Random graphs ,Law of Large Numbers for random graphs
Textbooks: 1. Nataraj Dasgupta, *Practical Big Data Analytics*, Packt Publishing, 2018. Covers fundamentals of Big Data, data processing, analytics frameworks, and real-world applications. 2. R. Elangovan, P. Rajesh et al., *Big Data Analytics (UG & PG, Computer Science & IT Students)*, 2025. Comprehensive textbook aligned with academic syllabus including regression, clustering, and analytics techniques. 3. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, *An Introduction to Statistical Learning*, Springer. Widely used for statistical learning concepts such as regression, classification, and model evaluation. 4. Peter Bruce, Andrew Bruce, Peter Gedeck, *Practical Statistics for Data Scientists*, O'Reilly. Focuses on applied statistical methods like hypothesis testing, estimation, and data analysis. 5. Balusamy, Abirami, Kadry, Gandomi, *Big Data: Concepts, Technology and Architecture*, Wiley. Covers Big Data technologies, architecture, and analytical frameworks in detail.
Reference Books: 1. Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer. Best for: Statistical learning theory, bias-variance, regression, clustering, high-dimensional data. 2. Jure Leskovec, Anand Rajaraman, Jeffrey Ullman, *Mining of Massive Datasets*, Cambridge University Press. Best for: Big Data concepts, clustering, large-scale data analytics, real-world applications. 3. Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer. Best for: Statistical inference, probabilistic models, multivariate analysis, machine learning foundations. NPTEL Course: Statistical Foundation for Big Data Analysis, By Prof. Arindam Banerjee | IIT Kharagpur https://onlinecourses.nptel.ac.in/noc26_cs09/preview