Doctor of Philosophy

Syllabus Proposed to be implemented from 2026 onward

Department of Data Science and Analytics

School of Mathematics, Statistics and Computational Sciences

Central University of Rajasthan

NH-8 Jaipur- Ajmer Highway, Bandarsindri

Kishangarh -305817

District-Ajmer, Rajasthan

Website: www.curaj.ac.in

Doctor of Philosophy

PROGRAM OBJECTIVES

Data Science is a highly multi-disciplinary subject. Therefore, the program Ph.D. in the Department of Data Science and Analytics aims to acquainted students with broad knowledge in data science and specific knowledge relevant to their own research interests, including theories and methods of intervention. The broad objectives of the Ph.D. program are listed below.

The students should:

1. have knowledge and skills that makes them proficient in their chosen area of specialization to conduct independent research and analysis.
2. have mastery of research methodology, analytical and methodological skills, advanced statistics, machine learning, data mining, etc.
3. have ability to design and conduct original and significant contributions in their area of specialization
4. have ability to communicate the results of their research in a clear and effective manner
5. have ability to teach and provide valuable educational experience to students in academic settings.

PROGRAM OUTCOMES

Upon getting acquainted with the various research method and recent trends/development in the field of data science, the objectives of the Ph.D. program will be achieved by ensuring the following outcomes.

1. Knowledge of the most advanced research in their area of specialization
2. Students will have broad insight, understanding and intuition of the whole process line of extracting knowledge from data
3. In-depth understanding the current state of the art in the individual research area and the ability to make substantial research contribution in the topic of their specialization
4. Communicate effectively in a variety of professional contexts.

PH.D. COURSE WORK STRUCTURE

The Ph.D. requirements in the Department of Data Science and Analytics have a limited amount of course work, essentially to prepare the student to carry out research, and to develop adequate breadth in the subject area. The course work shall be treated as prerequisite for Ph.D. preparation. For the coursework, the Department of Data Science and Analytics offers Research Methodology and Algorithms for Data Science as compulsory papers. Every student enrolled for Ph.D. in the Dept. of Data Science & Analytics need to register for at least three courses out of which two will be compulsory courses namely Research Methodology (Couse Code: DSA 701) and Algorithms for Data Science (Course Code: DSA 702) and one elective course from the list provided below. Furthermore, a Ph.D. student is encouraged to register / audit Ph.D. level courses in other departments also. The total number of credits for each student shall be not less than 12.

CENTRAL UNIVERSITY OF RAJASTHAN
Syllabus
Doctor of Philosophy

Ph.D. Course Work Structure

S. No.	Course Code	Course Title	Type of Course (C/DE)	Hours Per Week			Credits
				Lecture	Tutorial	Practical	
1	8.0DSA01	Research Methodology	C	3	0	2	4
2	8.0DSA02	Algorithms for Data Science	C	3	0	2	4
3	As per SWAYAM/ NPTEL	Research and Publication Ethics (RPE)(SWAYAM/NPTEL)	C	2	0	0	2
4	---	Pedagogy for Higher Education	C	3	0	0	3
5	8.0DSAXX	Discipline Elective	DE	3	0	2	4
Total				14	0	6	17

C: Core Course**, DE:** Discipline Elective

List of Discipline Electives

S. No.	Course Code	Course Title	Type of Course	Hours Per Week			Credits
				Lecture	Tutorial	Practical	
1	8.0DSA31	Machine Learning Revised: Applied Machine Learning	DE	3	0	2	4
2	8.0DSA32	Data Mining Revised: Data Mining and Knowledge Discovery	DE	3	0	2	4
3	8.0DSA33	Combinatorial Optimization	DE	3	0	2	4
4	8.0DSA34	Deep Learning	DE	3	0	2	4
5	8.0DSA35	Applied Natural Language Processing	DE	3	0	2	4
6	8.0DSA36	Advanced Statistical Methods Revised: Statistical Methods for Data Analysis and Modelling	DE	3	0	2	4
7	8.0DSA37	Data Pre-processing and Visualization	DE	3	0	2	4
8	8.0DSA38	New course: Scientific Writing and Computational Research	DE	2	0	2	4

Core Course

Programme: PhD in Data Science & Analytics Analytics

<table>
<tr><td colspan="4">Course: RESEARCH METHODOLOGY

Code: 8.0DSA01</td></tr>
<tr><td colspan="2"><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td colspan="2">Theory: 3
Tutorial: 0
Practical:2</td><td>End Semester Examination: 100 Marks</td><td>4</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Course Prerequisites: No</td></tr>
<tr><td colspan="4">Course Objectives:
● Provide the student with an overview of the research and research methodologies in CS.
● To make the students develop analytical thinking and data interpretation capability.
● To make the students learn the art of literature review and to focus on a research problem using scientific methods.
● They should have a good understanding of the process to choose suitable method(s) for the investigations.
● To learn how to report the results of research findings and write technical articles/research papers.</td></tr>
<tr><td colspan="4">Course Outcomes:</td></tr>
<tr><td colspan="4">Course Content:</td></tr>
<tr><td>Unit-I</td><td colspan="2">Introduction to research, Research ethic, Research methods in computer science, Analytical vs. Empirical methods, Quantitative, Qualitative, and Mixed methods, Case studies, Ethnography, Exposit facto, Survey.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-II</td><td colspan="2">Basic steps for doing research, Reviewing the literature, Formulation of research problem, Scope and objectives of research problem, Identifying variables, Constructing hypotheses. Guidelines for design of experiments, Error analysis and accuracy.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-III</td><td colspan="2">Data collection techniques, Analysis and interpretation of quantitative data, Descriptive statistics, Probability, Random variables, Sampling distribution and probability distribution, Hypothesis testing. Tests of significance, test of difference of mean and proportions, t-tests, ANOVA, Chi-square tests, Regression analysis, Multivariate analysis

Qualitative methods, Study designs, Elements and methods, The nature and types of qualitative research.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-IV</td><td colspan="2">Significance of literature review, Case Studies, Principles of report writing, purpose, nature and evaluation, structure and components of research report,</td><td>7T+4P (15 Hours)</td></tr>
</table>

	Format, Presentation of preliminary, Review process, Review guidelines, Validity threats, Review decisions.	

Text Book:

1. Michael P. Marder, *Research Methods for Science*, Cambridge University Press, 1st Edition, 2011.

Reference Books:

1. R. Panneerselvam, *Research Methodology*, PHI Learning, 1st Edition, 2004.
2. John W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, SAGE Publications, 4th Edition, 2014.
3. Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Elsevier, 1st Edition, 2010.
4. Robert A. Day and Barbara Gastel, *How to Write and Publish a Scientific Paper*, Cambridge University Press, 5th Edition, 1992.
5. Wayne C. Booth, Gregory G. Colomb, Joseph M. Williams, Joseph Bizup, and William T. FitzGerald, *The Craft of Research*, The University of Chicago Press, 4th Edition, 2016.

Course: ALGORITHMS FOR DATA SCIENCE

<table>
<tr><td colspan="3">Code: 8.0DSA02</td></tr>
<tr><td colspan="1"><u>TEACHING SCHEME:</u></td><td colspan="1"><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTE D:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical:2</td><td>End Semester Examination: 100 Marks</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Prerequisites: Basic knowledge in programming, probability, calculus and linear algebra.</td></tr>
<tr><td colspan="3">Course Objectives:
● To provide the students a solid background in algorithms.
● To enable the students to design algorithms to meet functional requirements and the target complexity bounds in terms of time and space complexity.
● To aid the students to develop data structure techniques for several aspects of programming.
● To enable the students to design algorithms to evaluate their actual performance compared to prospects from analysis.
● Implement and know the applications of algorithms for sorting, pattern matching etc</td></tr>
<tr><td colspan="3">Course Outcomes:</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Algorithm Analysis: Computational Tractability, Asymptotic Order of growth, Solving recurrence relations, Big O, Big Ω, Big Θ, Time and space complexity analysis, Master Theorem, Complexity Classes: P, NP, NP-Completeness.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Searching Algorithms: Linear Search, Binary Search with their analyses

Sorting Algorithms: Insertion Sort, Bubble Sort, Selection Sort, Shell Sort, QuickSort, MergeSort, HeapSort with their analyses</td><td>7T+4Ps (15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Design Paradigms: Divide and Conquer, Greedy, Dynamic Programming. Basic graph algorithms (BFS, DFS, strongly connected components), Minimum Spanning Trees (Prim and Kruskal), Single source Shortest path and all-pair shortest path, network flow</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Advanced data structures: Lists (Singly and doubly linked lists, Circular lists), Queues (Linear queues, circular queues, and priority queues, Double-ended queues), Binary Search Trees, Red-Black Trees, AVL trees, B-trees and B+-trees), Hashing (Hash functions, Simple hash functions, Dynamic resizing of hash tables), Heaps (Binary heap, min-heap, max-heap, Priority queues and their application, Disjoint Set Data Structures</td><td>7T+4P (15 Hours)</td></tr>
</table>

Text Books:

1. T. H. Cormen, C. E. Leiserson, R. L Rivest., and C. Stein, *Introduction to Algorithms*, MIT Press, 3rd Edition, 2009.
2. J. Kleinberg and E. Tardos, Algorithm Design, Pearson publication, 1st Edition, 2021.

Reference Books:

1. A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *The Design and Analysis of Algorithms*, Pearson, 1st Edition, 2009.
2. Stephen Marsland, *Machine Learning: An Algorithmic Perspective*, CRC Press, 1st Edition, 2009.

Discipline Elective

Discipline Elective Courses

<table>
<tr><td colspan="4">Course: APPLIED MACHINE LEARNING

Code: 8.0DSA31</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td colspan="2"><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical:2</td><td colspan="2">End Semester Examination: 100 Marks</td><td>4</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Course Prerequisites: Required background knowledge on probability, linear algebra, and should have good programming skills. The programming environment used in the practical will be Python</td></tr>
<tr><td colspan="4">Course Objectives:
To assist students to understand, and practice machine learning techniques.
To help the students to gain experience in developing and experimenting with real systems in a realistic environment using machine learning approaches.
Introduce several libraries and data sets that are publicly available, and will be used to illustrate the application of machine learning algorithms.
To develop the programming skills that will help students in building intelligent, adaptive artifacts.</td></tr>
<tr><td colspan="4">Course Outcomes:</td></tr>
<tr><td colspan="4">Course Content:</td></tr>
<tr><td>Unit-I</td><td colspan="2">Introduction to Machine Learning: Goals and Applications of Machine Learning (Supervised, Unsupervised, Semi-supervised Learning, Applications in Healthcare, Finance, Marketing, Robotics, Ethical considerations in Machine Learning, Challenges in Machine Learning).

Different Forms of Learning: Supervised Learning: Regression, Classification, Semi-supervised and Self-supervised Learning.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-II</td><td colspan="2">Regression and Classification Methods: Regression Analysis (Linear Regression, Ridge Regression, Lasso Regression, Bayesian Regression, Regression with Basis Functions).

Classification Methods: Linear Discriminant Analysis (LDA) (Assumptions, Dimensionality Reduction, Comparison with PCA), Logistic Regression (Sigmoid Function, Decision Boundaries, Regularization), Support Vector Machine (SVM) (Kernel Trick, Hyperplane, Support Vectors).</td><td>7T+4P (15 Hours)</td></tr>
</table>

Unit-III	Decision Trees: CART, Gini Impurity, Information Gain, Pruning, Random Forests, Decision Tree Ensembles, K-Nearest Neighbors: Distance Metrics, Choice of K, Curse of Dimensionality. Probabilistic Learning: Hidden Markov Models, Bayesian Networks, Markov Random Fields, Conditional Random Fields, Gaussian Mixture Models	**7T+4P (15 Hours)**
Unit-IV	Ensemble Methods: Boosting (Adaboost, Gradient Boosting, Bagging, Random Forest). Dimensionality Reduction: Principal Component Analysis, Independent Component Analysis, Multidimensional Scaling, and Manifold Learning. Reinforcement Learning: Multi-Agent Reinforcement Learning, Cooperation and Competition among Agents.	**7T+4P (15 Hours)**

Text Book:

1. Christopher Bishop, *Pattern Recognition and Machine Learning*, Springer, 1st Edition, 2006.

Reference Books:

1. Tom Mitchell, *Machine Learning*, McGraw-Hill Education, 1st Edition, 1997.
2. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2nd Edition, 2012.
3. Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, Elsevier, 3rd Edition, 2011.
4. Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, Springer, 1st Edition, 2001.
5. Stephen Marsland, *Machine Learning: An Algorithmic Perspective*, CRC Press, 2nd Edition, 2015.

Course: DATA MINING AND KNOWLEDGE DISCOVERY

Code: 8.0DSA32

TEACHING SCHEME:	**EXAMINATION SCHEME:**	**CREDITS ALLOTED:**
Theory: 3 **Tutorial: 0** **Practical:2**	**End Semester Examination: 100 Marks**	**4**

Course Prerequisites: NO

Course Objectives:

- Provide the student with an overview of the concepts of data mining and data pre-processing
- General understanding of various data mining tasks such as mining frequent patterns, classification and cluster analysis
- Knowledge of rough set theory and its application for classification and feature selection
- Develop knowledge and skills to evaluate the performance of different data learning algorithms
- Application of data mining techniques for various advanced applications such as web mining, text mining, temporal data mining etc.
- To help the students to gain experience in developing and experimenting with real systems in a realistic environment using data mining approaches.

Course Outcomes:

Course Content:

Unit-I	Data Mining Overview: Definition and importance of Data Mining, Evolution of Data Mining, Major categories of Data Mining tasks. Techniques in Data Mining: Supervised Learning vs. Unsupervised Learning, Classification, Clustering, Regression, Association Rule Mining. Issues and Challenges in Data Mining: Data Quality Issues, Scalability and Efficiency, Handling High-dimensional Data (Curse of Dimensionality). Data Mining Process: Handling Missing Data, Noise, and Outliers, Normalization and Transformation.	**7T+4P (15 Hours)**
Unit-II	Classification: Problem definition, General approach, Decision tree induction (CART, ID3 and C4.5 Algorithms, Pruning Techniques, Handling Continuous and Categorical Data in Decision Trees), Nearest neighbor classifiers, Bayesian classifiers, Support vector machine, Model evaluation (Accuracy, Precision, Recall, F1-Score, ROC Curve and AUC, Confusion Matrix and Error Analysis).	**7T+4P (15 Hours)**

	Association Rule Mining: Apriori Algorithm and its variants, Frequent Itemset Mining, Candidate Generation, Rule Pruning, Constraints on Support, Confidence, and Lift.	
Unit-III	Cluster analysis: Introduction, Importance of Clustering, Types of Clustering: Hard vs. Soft Clustering, Partitional Clustering vs. Hierarchical Clustering, Supervised vs. Unsupervised Clustering, Similarity and Distance Measures, Clustering with Different Distance Measures, Density-Based Clustering (DBSCAN), Center-Based Clustering Techniques (K-Means Algorithm and K-Medoids Algorithm), Hierarchical Clustering, Clustering Validation (Internal Validation and External Validation).	**7T+4P (15 Hours)**
Unit-IV	Advanced techniques: Web Mining (Introduction to Web Mining and its Categories, Web Content Mining, Web Structure Mining, Web Usage Mining), Text Mining (Unstructured Text Mining, Text Preprocessing, Text Clustering, Sentiment Analysis and Opinion Mining), Temporal Data Mining (Temporal Association Rules, Sequence Mining, Episode Discovery), Spatial Data Mining (Clustering, Classification, and Regression, DBSCAN for Spatial Data, K-Means with Spatial Constraints, Identifying Geospatial Patterns and Relationships).	**7T+4P (15 Hours)**

Text Book:

1. P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson Education India, 4th Edition, 2016.
2. Jiawei Han, Micheline Kamber, and Jian Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 3rd Edition, 2012.

Reference Books:

1. Arun K. Pujari, *Data Mining Techniques*, Universities Press, 1st Edition, 2016.
2. Michael A. Berry and Gordon S. Linoff, *Mastering Data Mining*, John Wiley & Sons, 1st Edition, 2000.
3. M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons, 1st Edition, 2011.

<table>
<tr><td colspan="4">Course: COMBINATORIAL OPTIMIZATION

Code: 8.0DSA32</td></tr>
<tr><td colspan="2"><u>TEACHING SCHEME:</u></td><td><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTE D:</u></td></tr>
<tr><td colspan="2">Theory: 3
Tutorial: 0
Practical:2</td><td>End Semester Examination: 100 Marks</td><td>4</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Course Prerequisites: NO</td></tr>
<tr><td colspan="4">Course Objectives:
● The students should be able to deal with mathematical tools for solving and analyzing combinatorial optimization problems.
● Develop knowledge and skills in linear and integer programming (LP&IP) and their applications
● Knowledge and skills in the area of general nonlinear and convex optimization
● Knowledge and implementation of discrete optimization problems and their algorithms
● To enable students to formulate combinatorial optimization problems as mathematical models in different domains of computing and data science</td></tr>
<tr><td colspan="4">Course Outcomes:</td></tr>
<tr><td colspan="4">Course Content:</td></tr>
<tr><td>Unit-I</td><td colspan="2">Linear Programming: Linear programming problem, Simplex method, Revised Simplex method, Duality, Dual Simplex, Interior Point Method.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-II</td><td colspan="2">Combinatorial Optimization Problems: Transportation problem, Assignment problem, Shortest path problem, Knapsack problem, Local search, Max-flow Min-cost problem.

Dynamic Programming, Branch and Bound.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-III</td><td colspan="2">Nonlinear Unconstrained Optimization: Local and global optimal, Convex optimization, Optimality conditions and gradient methods, Line searches and Newton's method, Conjugate gradient methods, Others recent advancement of gradient search.</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-IV</td><td colspan="2">Nonlinear Programming: Method to solve non-linear programming problem, Constraints optimization with inequality constraints, Quadratic programming problem – Kuhn tucker conditions, Wolfe’s modified simplex method, Duality in NLP.</td><td>7T+4P (15 Hours)</td></tr>
</table>

	Combinatorial Optimization Problems in Computer vision, social networks, cyber physical systems, Big Data analytics. Selected topics from research papers.	

Text Book:

1. Christos H. Papadimitriou, *Combinatorial Optimization: Algorithms and Complexity*, Dover Publications, 1st Edition, 1998.

Reference Books:

1. David G. Luenberger and Yinyu Ye, *Linear and Nonlinear Programming*, Springer, 4th Edition, 2016.
2. Bernhard Korte and Jens Vygen, *Combinatorial Optimization: Theory and Algorithms*, Springer, 5th Edition, 2018.
3. W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver, *Combinatorial Optimization*, Wiley-Interscience, 1st Edition, 1997.

Course: DEEP LEARNING **Code: 8.0DSA34**		
<u>TEACHING SCHEME:</u>	**<u>EXAMINATION SCHEME:</u>**	**<u>CREDITS ALLOTE D:</u>**
Theory: 3 **Tutorial: 0** **Practical:2**	**End Semester Examination: 100 Marks**	**4**
Course Prerequisites: NO		
Course Objectives: ● Introduce state-of-the-art deep learning algorithms, the problem settings, and their applications to solve real world problems. ● Learn to design neural network architectures and training procedures. ● Introduce students to deep learning platforms and software libraries		
Course Outcomes:		
Course Content:		

Unit-I	Introduction to Deep learning: Neuroscience inspiration, Neural Network Basics, Artificial Neural Network, Activation function, Cost functions, Backpropagation, Regularization, Feedforward Neural networks, Multi-layer neural network, Gradient descent and the backpropagation algorithm, Nesterov Accelerated Gradient descent, Adam, ReLU Heuristics for avoiding bad local minima, Dropout, Heuristics for faster training	**7T+4P (15 Hours)**
Unit-II	Convolutional Neural Networks: Architectures, convolution/pooling layers, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet. Recurrent Neural Networks: Long short-term memory (LSTM),Gated recurrent unit (GRU), Encoder Decoder architectures	**7T+4P (15 Hours)**
Unit-III	Deep Unsupervised Learning: Autoencoders (standard, sparse, denoising, contractive, etc.), Variational Autoencoders, Adversarial Generative Networks, Autoencoder and DBM	**7T+4P (15 Hours)**
Unit-IV	Attention and memory models, Dynamic memory networks. Applications of Deep Learning: Computer Vision, Natural Language Processing, Speech recognition, Large Language Models and Generative AI etc.	**7T+4P (15 Hours)**
Text Book:		

1. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, The MIT Press, 1st Edition, 2015.

Reference Books:

1. B. Yegnanarayana, *Artificial Neural Networks*, PHI Learning, 1st Edition, 2009.
2. Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media, 1st Edition, 2017.

<table>
<tr><td colspan="3">Course: APPLIED NATURAL LANGUAGE PROCESSING

Code: 8.0DSA35</td></tr>
<tr><td><u>TEACHING SCHEME:</u></td><td colspan="1"><u>EXAMINATION SCHEME:</u></td><td><u>CREDITS ALLOTED:</u></td></tr>
<tr><td>Theory: 3
Tutorial: 0
Practical:2</td><td>End Semester Examination: 100 Marks</td><td>4</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="3">Course Prerequisites: NO</td></tr>
<tr><td colspan="3">Course Objectives:
This course provides a thorough introduction to the essential components of practical Natural Language Processing (NLP) systems.
Understand and describe the evaluation of NLP systems.
Understand deep learning techniques used in NLP and apply them to solve machine translation and conversation problems.
Learn about major NLP issues and Identify possible future areas of NLP research.</td></tr>
<tr><td colspan="3">Course Outcomes:</td></tr>
<tr><td colspan="3">Course Content:</td></tr>
<tr><td>Unit-I</td><td>Introduction: Natural Language Processing (NLP), History of NLP, Neural Networks for NLP

Regular Expressions, Text Normalization, Edit Distance, N-gram Language Models

Naive Bayes and Sentiment Classification, Logistic Regression</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-II</td><td>Vector Space Model - word vectors, GloVe/Word2Vec model, word embedding, Text Classification, Clustering, and Summarization</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-III</td><td>Neural Networks and Neural Language Models, Sequence Processing with Recurrent Network, Encoder-Decoder Models, Attention and Contextual Embeddings

Statistical Machine Translation, Neural Machine Translation</td><td>7T+4P (15 Hours)</td></tr>
<tr><td>Unit-IV</td><td>Information Retrieval tasks using Neural Networks- Learn to Rank, Understanding Phrases, analogies</td><td>7T+4P (15 Hours)</td></tr>
</table>

	Applications – Sentiment Analysis, Spam Detection, Resume Mining, Conversation Modeling, Chat-bots, dialog agents, Question Processing	

Text Book:

1. Dan Jurafsky and James H. Martin, *Speech and Language Processing*, Pearson, 2nd Edition, 2009.

Reference Books:

1. Jacob Eisenstein, *Introduction to Natural Language Processing*, The MIT Press, 1st Edition, 2019.
2. Yoav Goldberg, *Neural Network Methods for Natural Language Processing*, Morgan and Claypool Publishers, 1st Edition, 2017.
3. Jason Brownlee, *Deep Learning for Natural Language Processing: Develop Deep Learning Models for Natural Language in Python*, Machine Learning Mastery, 1st Edition, 2019.
4. Steven Bird, Ewan Klein, and Edward Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, 1st Edition, 2009.

Course: STATISTICAL METHODS FOR DATA ANALYSIS AND MODELING Code: 8.0DSA36		
TEACHING SCHEME:	**EXAMINATION SCHEME:**	**CREDITS ALLOTED:**
Theory: 3 **Tutorial: 0** **Practical:2**	**End Semester Examination: 100 Marks**	**4**
Course Prerequisites: NO		
Course Objectives: ● To enable the students to carry out data analysis from the perspective of understanding statistics. ● To aid the students in understanding the types of questions that the statistical method addresses; ● The students should be able to employ data to make evidence-based decisions.		
Course Outcomes:		
Course Content:		

Unit-I	Estimation: Unbiasedness (Rao-Blackwell and Lehmann-Scheffé Theorem), Consistency (Weak/strong consistency, Law of Large Numbers, Central Limit Theorem), UMVUE (Fisher Information, Cramer-Rao Bound, properties of UMVUE), Maximum Likelihood Estimation (Definition, asymptotic properties, Fisher Information, Cramer-Rao Bound), Bayesian Estimation, Method of Moments, Robust Estimation, Multivariate Maximum Likelihood Estimation, Bayesian Estimation.	**7T+4P (15 Hours)**
Unit-II	Test of Hypotheses: Types of Errors (Type I, Type II errors, Power of a test, sample size determination), Test Statistic (t-test, z-test, chi-square test, F-test, large sample approximations), Parametric Tests for Means & Variances (One-sample, two-sample t-test, F-test, Levene's test, Bartlett's test).	**7T+4P (15 Hours)**
Unit-III	Gauss-Markov Model & ANOVA: Gauss-Markov Model (BLUE estimators, Generalized Least Squares (GLS)), Least Squares Estimators (OLS estimation, efficiency, assumptions (linearity, homoscedasticity), Analysis of Variance (ANOVA) (One-way, two-way ANOVA, F-statistic, post-hoc tests),	**7T+4P (15 Hours)**
Unit-IV	Regression: Multiple Linear Regression (OLS estimation, model selection (Adjusted R^2, AIC, BIC), regression diagnostics), Forward, Backward & Stepwise Regression (Predictor selection, criteria for entry/exit (p-value,	**7T+4P (15 Hours)**

	AIC, BIC)), Logistic Regression (Binary and multinomial logistic regression, model diagnostics, regularization (Lasso, Ridge)),	

TEXT BOOK:

1. Kishor S. Trivedi, *Probability and Statistics with Reliability, Queueing and Computer Science Applications*, John Wiley & Sons, 2nd Edition, 2001.

REFERENCE BOOKS:

1. P. J. Bickel and K. A. Docksum, *Mathematical Statistics*, Prentice Hall, 2nd Edition, 2000.
2. Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Academic Press, 5th Edition, 2014.

Course: DATA PRE-PROCESSING AND VISUALIZATION **Code: 8.0DSA37**		
<u>TEACHING SCHEME:</u>	**<u>EXAMINATION SCHEME:</u>**	**<u>CREDITS ALLOTED:</u>**
Theory: 3 **Tutorial: 0** **Practical:2**	**End Semester Examination: 100 Marks**	**4**
Course Prerequisites: Every student should have good programming skills. The programming environment to be used will be Python and R.		

Course Objectives:

- To make students understand what data preprocessing is and why it is needed as part of an overall data science and machine learning methodology
- To enable the students to review and understand data quality issues and how to address them
- To aid students in cleansing and transforming your data
- To make students be able to summarize their data by using data visualization.

Course Outcomes:

Course Content:

Unit-I	Introduction to Data Preprocessing: Meaning of data preprocessing, Data Preparation, Dirty data, Structuring Data, Data Quality	**7T+4P (15 Hours)**
Unit-II	Data Visualization: Introduction to Data Visualization, Basic principles, ideas and tools for data visualization, Exercise: create your own visualization of a complex dataset. Exploratory Data Analysis, Creating a Histogram, Box Plots, Bar Graphs, Other Graphs	**7T+4P (15 Hours)**
Unit-III	Data Preparation- Data Integration, Data Cleaning, Data Normalization, Data Transformation, Discretization Dealing with Missing Data, Dealing with Noisy Data Data Reduction-Dimensionality Reduction, Data Sampling, Binning and Reduction of Cardinality.	**7T+4P (15 Hours)**

Unit-IV	Introduction to Data Mining Software Package Including Data Preparation and Reduction: KEEL Large scale applications from signal processing, banking, marketing, etc.	**7T+4P (15 Hours)**

TEXT BOOK

1. Salvador García, Julian Luengo, and Francisco Herrera, *Data Preprocessing in Data Mining*, Springer Nature, 1st Edition, 2015.

REFERENCE BOOKS:

1. Cathy O'Neil and Rachel Schutt, *Doing Data Science: Straight Talk from the Frontline*, O'Reilly Media, 1st Edition, 2014.
2. Kirk Borne, *Data Science for Business and Decision Making: Data Preprocessing and Visualization*, Elsevier, 1st Edition, 2020.
3. Ben Fry, *Visualizing Data*, O'Reilly Media, 1st Edition, 2007.

Course: Scientific Writing and Computational Research **Code: 8.0DSA38**		
<u>TEACHING SCHEME:</u>	**<u>EXAMINATION SCHEME:</u>**	**<u>CREDITS ALLOTED:</u>**
Theory: 3 **Tutorial: 0** **Practical:2**	**End Semester Examination: 100 Marks**	**4**

Course Objectives:

- To develop proficiency in preparing high-quality scientific and technical documents using LaTeX.
- To enable scholars to create research papers, theses, reports, presentations, and professional academic documents.
- To provide practical knowledge of MATLAB for numerical computation, data analysis, visualization, and algorithm prototyping.
- To familiarize students with MATLAB toolboxes relevant to Computer Science research, including optimization, signal processing, and image processing.
- To introduce integration of MATLAB with Python for advanced scientific computing and research applications.

Course Content:

Unit-I	Fundamentals of LaTeX: Introduction to LaTeX and Overleaf, document structure, text formatting, lists, tables, multi-column layouts, sections and chapters, customization of fonts and colors, table of contents, headers, footers, page numbering, and preparation of academic documents such as reports and CVs.	**7T+4P (15 Hours)**
Unit-II	LaTeX for Research Publications: Mathematical equations and symbols, figures and graphics, cross-referencing, bibliography management using BibTeX, citations and references, preparation of research papers and theses, and presentation design using Beamer.	**7T+4P (15 Hours)**
Unit-III	MATLAB Programming and Visualization: MATLAB environment, variables, matrices and arrays, scripts and functions, loops and conditional statements, file handling, numerical computations, data analysis, and 2D/3D plotting and visualization techniques.	**7T+4P (15 Hours)**
Unit-IV	MATLAB Applications in Research: Signal processing, image processing, optimization and curve fitting, statistical analysis, numerical methods,	**7T+4P (15 Hours)**

	introduction to machine learning tools, MATLAB-Python integration, and applications of MATLAB in Computer Science research.	

TEXT BOOK

1. Lamport, Leslie. LaTeX: A Document Preparation System: User's Guide and Reference Manual. 2nd ed., Addison-Wesley, 1994.
2. Moore, Holly. MATLAB for Engineers. 6th ed., Pearson, 2021.
3. Attaway, Stormy. MATLAB: A Practical Introduction to Programming and Problem Solving. 6th ed., Butterworth-Heinemann, 2023.

REFERENCE BOOKS:

1. Oetiker, Tobias, et al. *The Not So Short Introduction to LaTeX2ε*. Version 6.4, 2021.
2. Hanselman, Duane C., and Bruce L. Littlefield. *Mastering MATLAB*. Pearson Education, 2011.
3. Gonzalez, Rafael C., Richard E. Woods, and Steven L. Eddins. *Digital Image Processing Using MATLAB*. 3rd ed., Pearson Education, 2020.