Nagar Yuwak Shikshan Sanstha's

Yeshwantrao Chavan College of Engineering

(An Autonomous Institution affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)

(Accredited 'A++' Grade by NAAC with a score of 3.6)

Hingna Road, Wanadongri, Nagpur - 441 110

Master of Technology

SoE & Syllabus 2025

M.Tech in Data Science

Nagar Yuwak Shikshan Sanstha's
Yeshwantrao Chavan College of Engineering
(An Autonomous Institution affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)
M.TECH. SCHEME OF EXAMINATION 2025
Department of Computer Technology

SoE No.
25DS-101

M.Tech in Data Science

Sl. No.	Sem	Course Code	Course Title	T/P	Contact Hours L	T	P	Hrs.	Credits	% Weightage TA	ESE	ESE Duration Hrs
I SEMESTER												
1	1	25DS101	Probability and Statistics	T	3	0	0	3	3	20	80	3
2	1	25DS102	Mathematics for Data Science	T	3	0	0	3	3	20	80	3
3	1	25DS103	Data Mining and Warehousing	T	3	0	0	3	3	20	80	3
4	1	25DS104	Research Methodology	T	3	0	0	3	3	20	80	3
5	1	25DS105	**Lab :** Data Preparation and Mining	P	0	0	2	2	1	60	40	
6	1	25DS106	**Lab :** Data Visualization and Analytics	P	0	0	2	2	1	60	40	
7	1	25DS107	**Lab :** Python for Data Science Lab	P	0	0	2	2	1	60	40	
8	1		**Professional Elective-I**	T	3	0	0	3	3	20	80	3
9	1		**Professional Elective- II**	T	3	0	0	3	3	20	80	3
			Total		**18**	**0**	**6**	**24**	**21**			

List of Professional Electives-I

1	1	25DS111	**PE-I :** Image Computing
2	1	25DS112	**PE-I :** Distributed Systems
3	1	25DS113	**PE-I :** Graph Mining

List of Professional Elective- II

1	1	23DS121	**PE-II :** Natural Language Processing
2	1	23DS122	**PE-II :** Time Series and Forecasting
3	1	23DS123	**PE-II :** Classical Optimization

II SEMESTER												
1	2	25DS201	Fundamentals of Machine Learning and Deep Learning	T	3	0	0	3	3	20	80	3
2	2	25DS202	**Lab :** Fundamentals of Machine Learning and Deep Learning	P	0	0	2	2	1	60	40	
3	2	25DS203	Big Data Analytics	T	3	0	0	3	3	20	80	3
4	2	25DS204	**Lab :** Big Data Analytics	P	0	0	2	2	1	60	40	
5	2	25DS205	Data Modeling	T	3	0	0	3	3	20	80	3
7	2	25DS206	**Lab :** Open Source Tools for Data Analysis	P	0	0	2	2	1	60	40	
8	2	25DS207	**Lab :** Prompt Engineering for Data Analysis	P	0	0	2	2	1	60	40	
9	2		**Professional Elective-III**	T	3	0	0	3	3	20	80	3
10	2		**Professional Elective-IV**	T	3	0	0	3	3	20	80	3
			TOTAL		**15**	**0**	**8**	**23**	**19**			

List of Professional Elective- III

1	2	25DS211	**PE-III :** Computer Vision
2	2	25DS212	**PE-III :** Cloud Fundamentals for Data Science
3	2	25DS213	**PE-III :** Social Network Analysis

List of Professional Elective- IV

1	2	25DS221	**PE-IV :** Text Analytics
2	2	25DS222	**PE-IV :** Information Retrival and Recommendation
3	2	25DS223	**PE-IV :** Non Classical Optimization
4	2	25DS224	**PE-IV :** Social Media Analytics

III SEMESTER												
1	3	23DS301	Project Phase -I	P	0	0	16	16	10	100		
			TOTAL		**0**	**16**	**16**	**10**				

IV SEMESTER												
1	4	25DS401	Project Phase-II	P	0	0	24	24	18	60	40	
			TOTAL		**0**	**24**	**24**	**18**				

Grand Total of Credits				**33**	**0**	**54**	**87**	**68**				

		June,2025	**1.00**
Chairperson	**Dean (Acad. Matters)**	**Date of Release**	**Version**

Applicable for AY 2025-26 Onwards

I SEMESTER
25DS101: Probability and Statistics

Course Outcomes :
Upon successful completion of the course the students will be able to 1. Reveal the hidden meaning in the data by applying some basic statistical formulae and probability distribution concepts using the tool 'R' 2. Employ the sampling techniques to find the estimates and test its validity using hypotheses testing 3. Analyze and compare sample data to make inference about the population data. 4. Design and implement the predictive model using simple and multiple regression technique

Unit I:	(7 Hrs.)
Introduction: Grouping and displaying data to convey meaning: Raw data, arranging data, frequency distribution, Measures of central tendency and dispersion in frequency distribution: arithmetic mean, weighted mean, geometric mean, Median, mode, dispersion, ranges, Exploratory data analysis(EDA). Introduction to R Statistics	
Unit II:	(7 Hrs.)
Probability and Probability distribution: Basic terminology in probability, probability rules, Probabilities under conditions of statistical independence, probabilities under conditions of statistical dependence. Probability distribution: What is probability distribution, random variables, use of expected value in decision making, and various distributions	
Unit III:	(6 Hrs.)
Sampling and Sampling Distribution and Estimation: Introduction to sampling, random sampling, Introduction to sampling distribution. Estimation: Introduction, Point estimates, Interval estimates and confidence interval, interval estimates using t distribution, determining the sample size in estimations	
Unit IV:	(7 Hrs.)
Testing Hypothesis: One sample test, Two sample tests: Introduction, testing hypothesis, hypothesis testing of means when the population standard deviation is known, measuring power of hypothesis, hypothesis testing of proportions, HT when standard deviation is not known, hypothesis testing for means and proportions, test for difference between means for various sample sizes	
Unit V:	(6 Hrs.)
Chi-square and analysis of Variance: Introduction, chi-square as a test of independence, chi-square as a test of goodness of fit: testing the appropriateness of a distribution, analysis of variance, inference about a population variance, Inference about two population variance	
Unit VI:	(6 Hrs.)
Simple Regression and Correlation and Multiple Regression and Modeling: Estimation using regression line, correlation analysis, making inference about population parameters, multiple regression and correlation analysis, finding the multiple regression equation, making inference about population parameters, modeling techniques	
Total Lecture	39 Hours

Textbooks:	
1.	Statistics for Management", Richard I. Levin & David S. Rubin, 7th Edition, Pearson Education.

Reference Books:	
1.	"Practical Statistics for Data Scientists, 50 Essential Concepts", Peter Bruce & Andrew Bruce, O'Reilly Media
2.	"An Introduction to Statistical Learning with Applications in R", Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani, Springer Press

MOOCs Links and additional reading, learning, video material	
1.	https://nptel.ac.in/courses/106106179
2.	https://www.youtube.com/watch?v=wrIvuzi56oQ

I Semester
25DS102: Mathematics for Data Science

Course Outcomes :

Upon successful completion of the course the students will be able to

1. Acquire knowledge on various mathematical concepts of Linear algebra to be used in Data Science.
2. Acquire the concepts of vector calculus.
3. Solve the various problems using optimization.
4. Solve data science problems through a guided approach.

Unit I:	(7 Hrs.)
Linear algebra: Systems of Linear equations, Solving Systems of Linear equations, Vector Spaces, Linear Independence, Basis and Rank linear mappings, AffineSpaces	
Unit II:	(7Hrs.)
Analytic Geometry: Norms, Inner Products, Lengths, and Distances Angles and Orthogonality, Orthonormal Basis, Orthogonal Complement, Inner Product of Functions, Orthogonal Projections, Rotations	
Unit III:	(6 Hrs.)
Matrix Decompositions: Determinant and Trace, Eigenvalues and Eigenvectors, Cholesky Decomposition, Eigendecomposition, and Diagonalization, Singular Value Decomposition, MatrixApproximation, and Matrix Phylogeny	
Unit IV:	(7Hrs.)
Vector Calculus: Differentiation of Univariate Functions, Partial Differentiation, and Gradients, Gradients Vector-Valued Functions, Gradient of Matrices, Useful Identities for Computing Gradients, Backpropagation, and Automatic Differentiation, Higher-Order Derivatives, Linearization and Multivariate Taylor Series	
Unit V:	(6 Hrs.)
Optimization in Data Science: Basics of optimization problem, Components of an Optimization Problem, Types of Optimization Problems, 1D optimization, Convex sets, Convex functions, and their properties unconstrained univariate optimization, nonlinear unconstrained multivariate optimization, Gradient (Steepest) Descent (OR) Learning Rule, Stochastic Gradient Descent, Method of the Conjugate Gradient	
Unit VI:	(6Hrs.)
Multivariate Optimization with Equality Constraints, Multivariate Optimization with Inequality Constraints, Solving Data Analysis Problems – A Guided Thought Process	
Total Lecture	**39 Hours**

Textbooks:	
1.	Mathematics for Machine Learning First M. P. Deisenroth, A. A. Faisal, and C. S. Ong Cambridge University Press
2.	Introduction to Optimization: Foundations and Fundamental Algorithms Niclas Andr´easson,Anton Evgrafov, and Michael Patriksson

Reference Books:	
1.	Mathematical Foundations of Data Analysis JEFF M. PHILLIPS
2.	Miller & Freund's Probability and Statistics for Engineers Eighth R. A. Johnson, I. Miller, and J. E.Freund Prentice Hall PTR

MOOCs Links and additional reading, learning, video material	
1.	https://nptel.ac.in/courses/106106179

I SEMESTER
25DS103 : Data Mining and Warehousing

Course Outcomes :
Upon successful completion of the course the students will be able to • Understand basic concepts of data mining and get an overview of various mining functionalities • Apply the techniques for supervised and unsupervised learning for knowledge extraction • Apply the concepts of frequent pattern mining and predictive data mining for knowledge extraction • Apply the concepts of data warehousing for designing multi dimensional data model and perform OLAP operations

Unit I:	(7 Hrs.)
Introduction to data mining: Process of data mining, Data Mining Functionalities, Classification of Data Mining systems, Data Mining Task primitives, Major issues in Data Mining, Applications of Data Mining	
Unit II:	(7 Hrs.)
Classification and Clustering: Classification: Introduction, decision tree, builing a decision tree – the tree induction algorithm, split algorithm based on information theory, gini index, over fitting and pruning, decision tree rules, naïve Bayes method. Types of data in cluster analysis, categorization of major clustering methods: Partitioning methods, Hierarchical methods, Applications of clustering.	
Unit III:	(7 Hrs.)
Mining Frequent Patterns and Association Rules: Market Basket Analysis, Frequent Item sets and Association rules, A priori Algorithm, Improving the efficiency of A priori, FP-growth algorithm.	
Unit IV:	(6 Hrs.)
Data mining using Prediction methods: Linear and nonlinear regression, Multivariate regression, Logistic Regression	
Unit V:	(6 Hrs.)
Introduction to data warehousing: Data warehousing components, Building a data warehouse, Multi- Dimensional Data Model, OLAP Operations in the Multi-Dimensional Model	
Unit VI:	(6 Hrs.)
Three Tier Data Warehouse Architecture, Data Warehouse Models, Schemas for Multi-dimensional data Model, OnlineAnalytical Processing (OLAP) - OLAP Vs OLTP, Integrated OLAM and OLAP Architecture	
Total Lecture	39 Hours

Textbooks:	
1.	Paul Raj Poonia, "Fundamentals of Data Warehousing", John Wiley & Sons
2.	Sam Anahony, "Data Warehousing in the real world: A practical guide for building decision support systems", John Wiley
3.	"Data Mining – Concepts and Techniques" Jiawei Han & Micheline Kamber Harcourt India
4.	"Data Mining Techniques" Arun K Pujari University Press.

Reference Books:	
1.	"Introduction to Data mining" Pang-ning Tan, Michael Steinbach, Vipin Kumar Pearson

MOOCs Links and additional reading, learning, video material	
1.	https://nptel.ac.in/courses/106105174

I SEMESTER
25DS104 : Research Methodology

Course Outcomes :
Upon successful completion of the course the students will be able to 1. **Explain** the foundations of research, including objectives, approaches, ethics, formulation of research questions, and literature review process. 2. **Apply** suitable techniques to define research problems, develop hypotheses, and design appropriate qualitative and quantitative research frameworks. 3. **Analyze and interpret** methods of data collection, sampling, inferential statistics, and research data using appropriate software, while ensuring ethical considerations 4. **Develop** the ability to prepare and present well-structured, ethical research reports using effective writing techniques, formatting tools, reference management, and plagiarism detection methods.

Unit I:	**6 Hrs.**
Research Foundations: Introduction to research, objective and importance of research, types of research, research approaches, significance of research, research methods vs methodology, research process, criteria of good research, ethics in research. Research formulation and literature review. Characteristics of good research question, literature review process.	
Unit II:	**6 Hrs.**
Defining Research Problem: Selecting the problem, necessity of defining the problem, technique involved in defining a problem, the planning process ,selection of a problem for research, formulation of the selected problems, hypothesis formation, measurement.	
Unit III:	**6 Hrs.**
Research Design: Meaning of research design, need for research design, features of a good design, important concepts relating to research design, different research designs, basic principles of experimental designs, qualitative and quantitative research.	
Unit IV:	**6 Hrs.**
Data Collection: Classification of data, methods of data collection, sampling fundamentals, sampling techniques procedure and methods, observation, surveys, inferential statistics, and interpretation of results. Ethical considerations in research	
Unit V:	**7 Hrs.**
Data Analysis and interpretation: Types of analysis, statistical techniques and choosing an appropriate statistical technique, hypothesis, hypotheses testing, data processing software, statistical inference, error analysis, interpretation of results and discussions.	
Unit VI:	**8 Hrs.**
Technical Writing and reporting of research: Significance of report writing, different steps in writing report, layout of the research report, types of reports, mechanics of writing a research report, precautions for writing research reports. **Use of tools / techniques for Research**: Methods to search required information effectively, reference management software like zotero/mendeley, software for paper formatting like latex/ms office, software for detection of plagiarism	
	Total Lecture **39 Hours**

Textbooks:	
1.	C. R. Kothari, "Research Methodology: Methods and Techniques", New Age International Publishers, ISBN:81-224-1522-9
2.	Fisher R. A, "Statistical Methods for Research Workers", Cosmo Publications, New Delhi, ISBN:81-307-0128-6 .

Reference Books:	
1.	Ranjit Kumar, Research Methodology: A Step-by-Step Guide for Beginners, Sage Publications.
2.	Montogomery D.C., "Design and Analysis of Experiments", (2001), John Wiley, ISBN: 0471260088.
3.	Robert P. Merges, Peter S. Menell, Mark A. Lemley, " Intellectual Property in New Technological Age".
4.	Creswell, John W. Research design: Qualitative, quantitative, and mixed methods approaches. Sage

I Semester
25DS105 – Data Preparation and Mining Lab

Course Outcome: After completion of the laboratory work, student will demonstrate the ability to

CO 1	Collect the data from heterogeneous sources
CO 2	Apply the data pre-processing techniques
CO 3	Perform the EDA on data
CO 4	Visualize the analysis drawn from the data

Syllabus:

Unit	Content	Hours
1	Data Exploration as a Process, The Nature of the World and Its Impact on Data Preparation, Defining data analysis problem	06
2	Data Preparation as a Process, Getting the Data—Basic Preparation, Sampling, Variability, and Confidence	07
3	Handling Nonnumerical Variables, Normalizing and Redistributing Variables, Replacing Missing and Empty Values	07
4	Series Variables, Preparing the Data Set, The Data Survey	07
5	Using Prepared Data: Exploratory data analysis	06
6	Using prepared data: Data Visualization. Case studies	07

Text Books:

SN	Title	Edition	Authors	Publisher
1	Data Preparation for Data Mining	NA	**Dorian Pyle**	Morgan Kaufmann Publishers
2	Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining	NA	**Glenn J. Myatt**	**Wiley–Blackwell**

Reference Books:

SN	Title	Edition	Authors	Publisher
1				

Website / Data sheet:

SN	Title
1	https://mostly.ai/blog/how-to-generate-synthetic-data

I SEMESTER
25DS106 : Data Visualization and Analytics Lab

Course Outcomes :
Upon successful completion of the course the students will be able to
1. Connect to and visualize data in Power BI 2. Build data model and get the insights from data. 3. Design compelling Power BI reports.

Lab Experiment List:

Expt. No	Name of Experiment
1	Introduction to Power BI and the different Power BI elements
2	Importing data into the Power BI from local data files and cloud servers
3	Clean, transform, and load data in Power BI
4	Create simple pre-defined models for visualization
5	Combine different visualization modes
6	Slice the dataset in Power BI
7	Matrices and tables in Power BI
8	Extract data relations and trends
9	Publish Power BI reports
10	Customizing the data analytics with Power BI and Power Automate

I SEMESTER
25DS107 : Python for Data Science Lab

Course Outcomes
Upon successful completion of the course the students will be able to 1: To create various classes and objects 2: To select the required framework and appropriate libraries to write a program in python 3: To develop an application using functionalities provided under various packages.

SN	Experiments based on
1	Write a program using object-oriented concept (class and object).
2	Write a program using Numpy.
3	Write a program using Pandas data frames and implement data frames related operations.
4	Write program on data manipulation. Analyzing type of data through file handling.
5	Program using: Matplotlib.
6	Write a program using Scipy.
7	Write a program using Scikit-learn.
8	Write a program using Tensorflow
9	Mini Project: Develop an application using the concept of data science

I Semester
25DS111 – PE I: Image Computing

Course Outcomes :
Upon successful completion of the course the students will be able to • Understand the need for image transforms different types of image transforms and their properties • Learn different techniques employed for the enhancement of images • Understand the need for image compression and to learn the spatial and frequency domain techniques of image compression • Learn different feature extraction techniques for image analysis and recognition

Unit I:	(7 Hrs.)
Overview of Digital Image Processing and Image Enhancement: A Simple Image Model, Sampling and Quantization, Basic Relationship Between Pixel, Basic gray level Transformation, Histogram Equalization, Histogram Processing, Local Enhancement, Image Subtraction, Image Averaging, Basics of Spatial Filtering, Smoothing Spatial Filtering, Sharpening Spatial Filters, Discrete Fourier Transformation, Fast Fourier Transformation, Fourier Properties, 2DFT, Inverse Fourier Transform, Filtering in Frequency Domain, Correspondence between Filtering in the Spatial and Frequency Domain, Smoothing Frequency Domain Filters, Sharpening Frequency Domain Filters, Homographic Filtering.	
Unit II:	(7 Hrs.)
Image Segmentation: Fundamentals, Point, line and edge detection, thresholding, Region Oriented Segmentation, Motion Based Segmentation.	
Unit III:	(7 Hrs.)
Morphing, Representation and Description: Introduction, Basic Morphological Algorithm, Chain Code, Polygonal Approximation, Signatures, Boundary Segments, Skeleton of a region, Boundary Descriptors, Shape Numbers, Fourier Descriptors, Regional Descriptors, Simple Descriptors, Topological Descriptors.	
Unit IV:	(6 Hrs.)
Model of the Image Degradation/Restoration Process: Noise Models, Restoration in the presence of Noise only-Spatial Filtering, Periodic Noise reduction by frequency domain filtering, Linear Position-Invariant Degradations, Estimation of Degradation Function, Inverse Filtering, Wiener Filtering, Constrained Least Square Filtering.	
Unit V:	(6 Hrs.)
Wavelets: Image Pyramids, Haar Transform, Multiresolution Expansions, Wavelet Transforms in1D, Fast wavelet Transform, wavelet packets.	

Unit VI:	(6 Hrs.)
Image Compression: Fundamentals of Image compression, coding redundancy, spatial and temporal redundancy, Irrelevant Information, Measuring Image Information, Fidelity criteria, Image compression models, compression standards, Basic compression methods, Huffman coding, colomb coding, arithmetic coding, LZW coding, runlength coding, Symbol based coding, Block transform coding, predictive coding.	
Total Lecture	**39 Hours**

Textbooks:			
1.	Digital Image Processing 3rd Edition	Rafael C. Gonzalez & Richard E. Woods	Pearson Education.
2.	Fundamental of Digital Image Processing	A. K. Jain	PHI.

Reference Books:		
1.	Digital Image Processing	Rosefield Kak
2.	Digital Image Processing	W. K. Pratt
3.		

MOOCs Links and additional reading, learning, video material	
1.	https://onlinecourses.nptel.ac.in/noc19_ee55/preview

I SEMESTER
25DS112 : PE I: Distributed Systems

Course Outcomes :
Upon successful completion of the course the students will be able to 1. To develop and apply knowledge of distributed systems techniques and methodologies. 2. To gain experience in the design and development of distributed systems and distributed systems applications. 3. To gain experience in the application of fundamental Computer Science methods and algorithms in the development of distributed systems and distributed systems applications. 4. To gain experience in the design and testing of a large software system, and to be able to communicate that design to others.

Unit I:	(7 Hrs.)
Characterization of Distributed Systems: Introduction, Examples of Distributed Systems, Resource Sharing and The Web, Challenges, System Models: Introduction, Architectural Models- Software Layers, System Architecture, Variations, Interface and Objects, Design Requirements for Distributed Architectures, Fundamental Models- Interaction Model, Failure Model, Security Model.	
Unit II:	(7 Hrs.)
Interprocess Communication: Introduction, The API for the Internet Protocols- The Characteristics of Interprocess communication, Sockets, UDP Datagram Communication, TCP Stream Communication; External Data Representation and Marshalling; Client Server Communication; Group Communication- IP Multicast- an implementation of group communication, Reliability and Ordering of Multicast.	
Unit III:	(7 Hrs.)
Distributed Objects and Remote Invocation: Introduction, Communication between Distributed Objects- Object Model, Distributed Object Model, Design Issues for RMI, Implementation of RMI, Distributed Garbage Collection; RPC, Events and Notifications, Case Study: JAVA RMI	
Unit IV:	(6 Hrs.)
Operating System Support: Introduction, The Operating System Layer, Protection, Processes and Threads – Address Space, Creation of a New Process, Threads.	
Unit V:	(6 Hrs.)
Distributed File Systems: Introduction, File Service Architecture; Peer-toPeer Systems: Introduction, Napster and its Legacy, Peer-to-Peer Middleware, Routing Overlays.	
Unit VI:	(6 Hrs.)
Coordination and Agreement: Introduction, Distributed Mutual Exclusion, Elections, Multicast Communication. Transactions& Replications: Introduction, System Model and Group Communication, Concurrency Control in Distributed Transactions, Distributed Dead Locks, Transaction Recovery; Replication- Introduction, Passive (Primary) Replication, Active Replication.	
Total Lecture	**39 Hours**

Textbooks:	
1.	Distributed Systems- Concepts and Design Fourth Edition George Coulouris, Jean Dollimore, Tim Kindberg, Pearson Publication

2.	Distributed Computing, Principles, Algorithms and Systems Ajay D Kshemkalyani, Mukesh Sighal, Cambridge
3.	Distributed Systems- Principles and Paradigms Andrew S Tanenbaum, Maarten Van Steen Pearson Publication

Reference Books:	
1.	Designing Distributed Systems: Patterns and Paradigms for Scalable, Reliable Services Bredan Burns Kindle eTextbook store.

MOOCs Links and additional reading, learning, video material	
1.	https://onlinecourses.nptel.ac.in/noc21_cs87/preview

I Semester
25DS113 : PE I: Time Series Data Analysis

Course Outcomes :
Upon successful completion of the course the students will be able to • Demonstrate an understanding of the concepts of time series analysis and models • Demonstrate an understanding of stationary process and ARMA models used for modeling and forecasting • Demonstrate an understanding of nonstationary and seasonal time series models • Demonstrate an understanding of multivariate time series • Demonstrate an understanding of State-space representations and estimation of time series models

Unit I:	(7 Hrs.)
Introduction: Examples of time series, Objectives of Time Series Analysis, Simple Time Series Models, Stationary models and autocorrelation functions Estimation and elimination of trend and seasonal components	
Unit II:	(7 Hrs.)
Stationary Process and ARMA Models: Basic properties and linear Processes Introduction to ARMA models, properties of sample mean and autocorrelation function, Forecasting stationary time series, ARMA (p, q) processes, ACF and PACF, and forecasting of ARMA processes, SpectralAnalysis	
Unit III:	(7 Hrs.)
Modeling and Forecasting with ARMA Processes: Preliminary estimation , Maximum likelihood estimation, Diagnostics, Forecasting, and order selection	
Unit IV:	(6 Hrs.)
Nonstationary and Seasonal Time Series Models: ARIMA model, Identification techniques, Unit roots in time series, Forecasting ARIMA models, Seasonal ARIMA models, Regression with ARMA errors	
Unit V:	(6 Hrs.)
Multivariate Time Series: Second-order properties of multivariate time series, Estimation of the mean and covariance, Multivariate ARMA processes, Best linear predictors of second-order random vectors, Modeling and forecasting	
Unit VI:	(6 Hrs.)
State-Space Models: State-space representations, The basic structure model, State-space representations of ARIMA models, The Kalman Recursions, Estimation for state-space models, and estimation of time series models	
Total Lecture	**39 Hours**

Textbooks:	
1.	Introduction to Time Series and Forecasting Second Brockwell, Peter J. and Davis, Richard A Springer-Verlag, New York
2.	

Reference Books:	
1.	Time Series Analysis: Forecasting and Control, Third Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. Prentice Hall, New Jersey.
2.	The Analysis of Time Series, Eighth Chatfield, C. Chapman and Hall, New York.
3.	

MOOCs Links and additional reading, learning, video material	
1.	TimeSeriesAnalysisand ItsApplications With R Examples, EZ-ThirdEdition

I SEMESTER
25DS121: PE II: Natural Language Processing

Course Outcomes :
Upon successful completion of the course the students will be able to 1. Model linguistic phenomena with formal grammars. 2. Design, implement and test algorithms for NLP problems 3. Apply NLP techniques to design real world NLP applications

Unit I:	(7 Hrs.)
Introduction to NLP: Computational Models of Language, Organization of NLP Systems, Natural Language Generation.	
Unit II:	(7 Hrs.)
Syntax: Linguistic Background, Elements of Simple Sentences, Parsing Techniques, Features and Augmented Grammars, Deterministic Parsing	
Unit III:	(7 Hrs.)
Semantic: Logical Form, Case Relations, Semantic Networks.	
Unit IV:	(6 Hrs.)
Context & World Knowledge: Knowledge Representation, Question, Answering Systems: Natural Language Generation, Typical NLP Systems and their Architectures, Cognitive Aspects of Natural Languages	
Unit V:	(6 Hrs.)
Indian Language Processing: Techniques of Machine Translation, Approaches to Machine Translation, Typical Case Studies in Indian Language Context	
Unit VI:	(6 Hrs.)
Introduction to Speech Processing: Word level Morphology and Computational Phonology; Basic Text to Speech; Introduction to HMMs and Speech Recognition, Part of Speech Tagging; Parsing with CFGs; Probabilistic Parsing. Representation of Meaning; Semantic Analysis; Lexical Semantics; Word Sense; Disambiguation; Discourse understanding; Indian language case studies	
Total Lecture	**39 Hours**

Textbooks:				
1.	"Natural Language Understanding",	First Edition	James Allen	Pearson Education.
2.	"Speech and Language Processing",	First Edition	Daniel Jurafsky and James H. Martin,	Prentice-Hall

Reference Books:	
1.	Foundations of Statistical Natural Language Processing First Edition Christopher Manning MIT Press, Cambridge
2.	Natural Language Processing Third Edition Akshar Bharathi, Vineet Chaitanya, Rajeev Sangal, – A Paninian Perspective", Prentice Hall
3.	Foundations of Computational Linguistics First Edition Ronald Hausser, Springer-Verlog,

MOOCs Links and additional reading, learning, video material	
1.	https://onlinecourses.nptel.ac.in/noc23_cs45/preview

II SEMESTER

25DS201: Fundamentals of Machine Learning and Deep Learning

Course Outcomes :
Upon successful completion of the course the students will be able to • Interpret machine learning techniques suitable for a given problem • Apply machine learning techniques to solve the problems • Compare machine learning techniques • Evaluate different machine learning techniques

Unit I:	(7 Hrs.)
Machine Learning: Introduction, Supervised algorithms, Unsupervised algorithms, Reinforcement, Bias variance trade-off, loss functions, experimentation and evaluation metrics.	
Unit II:	(7 Hrs.)
Supervised Machine Learning: Bayes learning, K- nearest neighbor learning, Linear regression, logistic regression, introduction to support vector machines, kernel functions	
Unit III:	(6Hrs.)
Unsupervised Machine Learning: Density estimation, Clustering, Dimensinality reduction, PCA	
Unit IV:	(7 Hrs.)
Artificial Neural Networks: Biological neural network, Artificial neural network, Hopfield network, Perceptron, Multilayer networks and Backpropagation algorithm	
Unit V:	(6 Hrs.)
Deep Learning: History of deep learning, perceptron learning algorithm, Multi-Layer Network and Optimization Technique, Dimension Reduction and Regularization, Convolutional Neural Networks: lenet, alexnet, zf-net, vggnet, googlenet, resnet, applications of convolutional neural networks	
Unit VI:	(6 Hrs.)
Recurrent Neural Networks: back propagation through time (bptt), vanishing and exploding gradients, LSTM encoder decoder models, attention mechanism, Applications of RNN	
Total Lecture	39 Hours

Textbooks:	
1.	"Introduction to Machine Learning"Third Ethem Alpaydin The MIT Press
2.	"Machine Learning"Second Tom M. Mitchell McGraw-Hill Education India Private Limited

Reference Books:

1.	Christopher M. Bishop, Pattern Recognition and Machine Learning. http:// research.microsoft.com/enus/um/people/cmbishop/prml/
2.	R. Sutton and A. Barto, An Introduction to Reinforcement Learning (http://webdocs.cs.ualberta.ca ~sutton/book/ebook/thebook.html)

MOOCs Links and additional reading, learning, video material	
1.	https://nptel.ac.in/courses/106106139
2.	https://onlinecourses.nptel.ac.in/noc20_cs62/preview

II Semester

25DS202 : Fundamentals of Machine Learning and Deep Learning Lab

Lab Experiment List:

Expt. No	Name of Experiment
1	Introduction to popular Machine Learning Datasets and Toolkits
2	Face Recognition using SVM
3	Practical applications of clustering
4	Experiments on supervised classification
5	Application of Classifiers
6	Sequence classification using HMM
7	Applications of CNN
8	Applications of RNN

II Semester
25DS203 : Big Data Analytics

Course Outcomes :
Upon successful completion of the course the students will be able to 1. Understand the characteristics of big data and concepts of Hadoop ecosystem 2. Understand the concepts of Scala programming 3. Apply Mapreduce programming model to process big data 4. Analyze Spark and its uses for big data processing 5. Design programs for big data applications using Hadoop components

Unit I:	**7 Hrs.)**
Introduction to Big data: Introduction – Big Data- Characteristics of Big Data – Big data management architecture- Examining Big Data Types – Big Data Technology Components -– Big data analytics –Big data analytics examples - Web Data Overview – Web Data in Action.	
Unit II:	(7Hrs.)
Hadoop : Introduction – History of Hadoop - Hadoop Ecosystem- Analyzing data with Hadoop - Hadoop Distributed File System- Design - HDFS concepts - Hadoop filesystem –Data flow – Hadoop I/ O - Data integrity – Serialization - Setting up a Hadoop cluster - Cluster specification - cluster setup and installation	
Unit III:	(7 Hrs.)
MapReduce: Introduction – Understanding Map, Reduce functions - Scaling out - Anatomy of a MapReduce Job Run - Failures – Shuffle and sort - Mapreduce types and formats - features – counters - sorting - Mapreduce Applications – Configuring and setting the environment - Unit test with MR unitlocal test	
Unit IV:	(6Hrs.)
Spark: – Installing spark – Spark applications, Jobs, Stages and Tasks –Resilient Distributed databasesAnatomy of a Spark Job Run – Spark on YARN- SCALA: Introduction- Classes and objects- Basic types and operators- built-in control structures- functions and closures- inheritance	
Unit V:	(6 Hrs.)
NoSQL Databases: Introduction to NoSQL- MongoDB: Introduction – Data types – Creating, Updating and deleing documents -Querying – Introduction to indexing – Capped collections. Hbase: Concepts - Hbase Vs RDBMS - Creating records- Accessing data – Updating and deleting data – Modifying dataexporting and importing data	
Unit VI:	(6Hrs.)
USE CASES: Call detail log analysis, Credit fraud alert, Weather forecast	
Total Lecture	**39 Hours**

Textbooks:	
1.	"Hadoop: The Definitive Guide "Third Edit on Tom White O'reily Media, 2012
2.	"Big Data Analytics" First Edition Seema Acharya, Subhasini Chellappan " Wiley 2015.

Reference Books:	
1.	"Taming the Big Data Tidal wave "Bill Franks (2012). John Wiley & Sons
2.	"Programming in Scala", Second Edition, Martin Odersky, Lex Spoon, Bill Venners (2010) Artima Press, California.
3.	"Professional NoSQL"Shashank Tiwari (2011). John Wiley & Sons

MOOCs Links and additional reading, learning, video material	
1.	https://onlinecourses.nptel.ac.in/noc20_cs92/preview
2.	https://onlinecourses.swayam2.ac.in/arp19_ap60/preview

II Semester
25DS204 : Big Data Analytics Lab

Course Outcome: After completion of the laboratory work, student will demonstrate the ability to

Course Outcomes :
Upon successful completion of the course the students will be able to 1. Understand hadoop and its ecosystes. 2. Implementation of HDFS 3. Apply MapReduce on various sets of data. 4. Understand basics of NoSQL Databases. 5. Apply databse functionalities on datasets.

Lab Experiment List:

Expt. No	Name of Experiment
1	Downloading and installing Hadoop; Understanding different Hadoop modes. Startup scripts, Configuration files.
2	Hadoop Implementation of file management tasks, such as Adding files and directories, Retrieving files and Deleting files
3	Implement of Matrix Multiplication with Hadoop Map Reduce
4	Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm
5	Implementation of K-means clustering using Map Reduce
6	Installation of Hive along with practice examples
7	Installation of HBase, Installing thrift along with Practice examples
8	Patrice importing and exporting data from various data bases .
9	Installation of MongoDB database.
10	Creation of database in MongoDB platform and apply various operations on it.

Link for Lab Mannual:
http://deccancollege.ac.in/MCALABMANUALS/BIGDATALABMANUAL.pdf

II Semester
25DS205 : Data Modeling

Course Outcomes :
Upon successful completion of the course the students will be able to 1. Understand the concepts of relational database modelling, multidimensional data modelling, unstructured data modelling 2. Apply the knowledge of database modelling concepts for structured data to create the database model 3. Apply the knowledge of database modelling concepts for un-structured data to create the database model 4. Analyze the data to find the suitable data modelling approach.

Unit I:	**(6 Hrs.)**
Introduction: Concepts of Data Modelling, Data Modelling Types, Data Model Standards, Business Requirements. Relational data base modeling concepts. Creation logical data model, creation physical data model, implementation of data models into databases. ER approach (subtypes and supertypes, Extensions and Alternatives), advanced Normalization concepts	
Unit II:	**(7 Hrs.)**
Multidimensional Data Model: OLAP and OLTP Concepts, Multidimensional Data Modelling, Concepts of facts, dimension, Types of facts and dimensions, types of schemas. Time dependant data, Data Cube Technology, Modelling for Data Warehouses and Data Marts	
Unit III:	**7 Hrs.)**
Enterprise Data Models and Data Management, aggregate data models, More details on data models, Relationships, Graphs databases, schemaless databases, Materialized views, modelling for data access. Data Models for GIS (Geographical Information System)	
Unit IV:	**(7 Hrs.)**
Modelling Unstructured data: Introduction to NoSQL databases, Basic Map Reduce, Partitioning and Combining, Composing Map Reduce Calculations, Key – Value databases. What is Key – Value Store?, Key – Value store features, transactions, structure of data, case studies based on actual data bases	
Unit V:	**(6 Hrs.)**
Document Databases: Introduction, Features, Consistency, Transactions, availability, Query Features, scaling, suitable use cases	
Unit VI:	**(6 Hrs.)**
Graph Databases: Introduction of the graph databases, features, consistency, Transactions, Availability, Query Features, suitable use cases	
Total Lecture	**39 Hours**

Textbooks:	
1.	Data Modelling Essentials 3rd Edition Graeme C. SimSion, Graham C. Witt MORGAN KAUFMANN PUB.
2.	Data Mining Concepts and Techniques Latest Jiawei Han, Micheline Kamber, Jian Pei MORGAN KAUFMANN PUB.
3.	NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence NA Sadalage, P. & Fowler Wiley Publications,1st Edition ,2019

Reference Books:	
1.	Fundamentals of Business Analytics Latest R. N. Prasad, Seema Acharya Wiley India

II Semester
25DS206: Open Source Tools for Data Analysis Lab

Course Outcomes :
Upon successful completion of the course the students will be able to 1. Connect to and visualize data in Power BI 2. Build data model and get the insights from data. 3. Design compelling Power BI reports.

Lab Experiment List:

Expt. No	Name of Experiment
1	Introduction to Power BI and the different Power BI elements
2	Importing data into the Power BI from local data files and cloud servers
3	Clean, transform, and load data in Power BI
4	Create simple pre-defined models for visualization
5	Combine different visualization modes
6	Slice the dataset in Power BI
7	Matrices and tables in Power BI
8	Extract data relations and trends
9	Publish Power BI reports
10	Customizing the data analytics with Power BI and Power Automate

II Semester
25DS207: Prompt Engineering for Data Analysis Lab

Course Outcomes :
Upon successful completion of the course the students will be able to
1. Increase productivity through chatGPT 2. Improve Critical Thinking 3. Learn The Fundamentals Of AI And NLP

Lab Experiment List:

Expt. No	Name of Experiment
1	Introduction to prompt Engineering for Data Analysis Python, Pandas, ChatGPT [Link:https://www.udemy.com/course/chatgptandpython/?couponCode=LETSLEARNNOWPP]
2	Prompts for general coding workflows
3	Prompts for data analysis workflows
4	Prompts for data visualization workflows
5	Prompts for machine learning workflows
6	Prompts for time series analysis workflows
7	Prompts for natural language processing workflows
8	Project Report

II Semester
25DS211 : PE III: Computer Vision

Course Outcomes :
Upon successful completion of the course the students will be able to • Identify basic concepts, terminology, theories, models and methods in the field of computer vision, • Describe basic methods of computer vision related to multi-scale representation, edge detection and detection of other primitives, stereo, motion and object recognition, • Suggest a design of a computer vision system for a specific problem

Unit I:	**(7 Hrs.)**
Introduction: Introduction to Human and Computer Vision, Image Registration algorithm	
Unit II:	**(6 Hrs.)**
Pattern Recognition Techniques: Statistical, Structural, Neural and Hybrid Techniques, Feature Extraction Techniques, Training and Classification	
Unit III:	**(7 Hrs.)**
Stereo Vision: Sensing 3D Shapes, How the 3rd dimension changes the problem, Stereo 3D description,3D Model, Matching	
Unit IV:	**(7 Hrs.)**
CBIR: Introduction, Content based image retrieval	
Unit V:	**(6 Hrs.)**
Virtual Reality: Introduction, basics of Virtual reality	
Unit VI:	**(6 Hrs.)**
Emerging CV applications: Recognition of characters, Fingerprint, Iris and Face	
Total Lecture	**39 Hours**

Textbooks:	
1.	Shapiro and G. Stockman, "Computer Vision", Prentice Hall
2.	David A. Forsyth, Jean Ponce, "Computer Vision", Prentice Hall
Reference Books:	
1.	Milan Sonka,Vaclav Hlavae, "Image Processing and Machine Vision" J.T. Tou and R. C. Gonzalez , "Pattern Recognition Principles"

MOOCs Links and additional reading, learning, video material	
1.	

II Semester
25DS212 : PE III: Cloud Fundamentals for Data Science

Course Outcomes :
Upon successful completion of the course the students will be able to • Characterize the distinctions between Infrastructure, Platform and Software as a Service (IaaS, PaaS,SaaS) abstractions; • Analyze the advantages and disadvantages of Public and Private Clouds. • Develop and deploy cloud application using popular cloud platforms. • Design Cloud security solutions

Unit I:	**(7 Hrs.)**
Introduction to Cloud Computing: Origins and Influences; Basic Concepts andTerminology; Goals and Benefits; Risks and Challenges. Fundamental Concepts and Models: Roles and Boundaries; Cloud Characteristics; CloudDelivery Models; Cloud Deployment Models	
Unit II:	**(6 Hrs.)**
Cloud Computing Technologies: Broadband Networks and Internet Architecture; Data Center Technology; Virtualization Technology; Web Technology; Multitenant Technology; Service Technology; Case study.	
Unit III:	**(7 Hrs.)**
Cloud Infrastructure Mechanisms: Logical Network Perimeter; Virtual Server; Cloud Storage Device; Cloud Usage Monitor; Resource Replication; Ready-made environment. **Specialized Cloud Mechanisms:** Automated Scaling Listener; Load Balancer; SLA Monitor;Pay-per-use Monitor; Audit Monitor	
Unit IV:	**(7 Hrs.)**
Cloud Management Mechanisms: Remote Administration System; Resource Management System; SLA Management System; Billing Management System. **Cloud Security:** Basic Terms and Concepts; Threat Agents; Cloud Security Threats; Additional considerations.	
Unit V:	**(6 Hrs.)**
Audit and compliance: Internal policy compliance, Givernance, Risk and Compliance (GRC), Regularity/External Compliance, Cloud Security Alliance, Auditing the cloud for Compliance, Security-as-a-Cloud.	

Unit VI:	**(6 Hrs.)**
Introduction to Hybrid Cloud: Hybrid Cloud Management, Managing the hybrid workloads, Development and deployment in Hybrid Cloud.	
Total Lecture	**39 Hours**

Textbooks:	
1.	Cloud Computing: Concepts, Technology &Architecture 2013 Thomas Erl, Ricardo Puttini, Zaigham Mahmood PHI
2.	Distributed and Cloud Computing 2012 Kai Hwang, Geoffrey C. Fox, Jack J Dongarra MK
3.	Grid and Cloud Computing, 2016 DharanipragadaJanakiram McGraw-Hill

Reference Books:	
1.	Cloud Computing: Theory and Practice Dan C. Marinescu MK
2.	Cloud Computing: Web-Based Applications That Change the Way You Work andCollaborate Online August 2008 Michael Miller, Que Publishing
3.	Cloud Computing- Principles andPradigms RajkumarBuyya, JamesBroberg, AndrzejGoscinski, Wiley
4.	Cloud Computing, A practical approach Anthony T.Velte, Toby J.Velte, Robert Elsenpeter TATAMcGRAW HILL
5.	Enterprise Cloud Computing- Technology, Architecture, Applications Gautam Shroff CAMBRIDGE

MOOCs Links and additional reading, learning, video material	
1.	

II Semester
25DS213 : PE III: Social Network Analysis

Course Outcomes :
Upon successful completion of the course the students will be able to • Develop semantic web related applications. • Represent knowledge using ontology. • Predict human behaviour in social web and related communities. • Visualize social networks.

Unit I:	(6 Hrs.)
INTRODUCTION:Introduction to Semantic Web: Limitations of current Web - Development of Semantic Web - Emergence of the Social Web - Social Network analysis: Development of Social Network Analysis - Key concepts and measures in network analysis - Electronic sources for network analysis: Electronic discussion networks, Blogs and online communities - Web-based networks - Applications of Social Network Analysis	
Unit II:	(7 Hrs.)
MODELLING, AGGREGATING: Ontology and their role in the Semantic Web: Ontology-based knowledge Representation - Ontology languages for the Semantic Web: Resource Description Framework - Web Ontology Language - Modelling and aggregating social network data	
Unit III:	(7 Hrs.)
KNOWLEDGE REPRESENTATION:State-of-the-art in network data representation - Ontological representation of social individuals - Ontological representation of social relationships - Aggregating and reasoning with social network data - Advanced representations	
Unit IV:	(7 Hrs.)
EXTRACTION AND MINING COMMUNITIES IN WEB SOCIAL NETWORKS: Extracting evolution of Web Community from a Series of Web Archive - Detecting communities in social networks - Definition of community - Evaluating communities - Methods for community detection and mining - Applications of community mining algorithms - Tools for detecting communities social network infrastructures and communities - Decentralized online social networks - Multi-Relational characterization of dynamic social network communities.	
Unit V:	(6 Hrs.)
PREDICTING HUMAN BEHAVIOUR AND PRIVACY ISSUES: Understanding and predicting human behaviour for social communities - User data management - Inference and Distribution - Enabling new human experiences - Reality mining - Context - Awareness - Privacy in online social networks - Trust in online environment - Trust models based on subjective logic - Trust network analysis - Trust transitivity analysis -	
Unit VI:	(7 Hrs.)
VISUALIZATION AND APPLICATIONS OF SOCIAL NETWORKS: Graph theory - Centrality - Clustering - Node-Edge Diagrams - Matrix representation - Visualizing online social networks, Visualizing social networks with matrix-based representations - Matrix and Node-Link Diagrams - Hybrid representations - Applications - Cover networks - Community welfare - Collaboration networks - Co-Citation networks.	
Total Lecture	39 Hours

Textbooks:				
1.	Social Networks and the Semantic Web	First Edition	Peter Mika	Springer 2007
2.	Handbook of Social Network Technologies and Applications	1st Edition	Borko Furht	Springer 2010

Reference Books:				
1.	Web Mining and Social Networking – Techniques and applications	First Edition	Guandong Xu	,Yanchun Zhang and Lin Li Springer, 2011.

MOOCs Links and additional reading, learning, video material	
1.	

II Semester
25DS221: PE IV: Text Analytics

Course Outcomes :
Upon successful completion of the course the students will be able to • Familiarize the learners with the concept of social media analytics and understand its significance. • Familiarize the learners with the tools of social media analytics. • Enable the learners to develop skills required for analyzing the effectiveness of social media for business purposes

Unit I:	(6 Hrs.)
Introduction to Social Media Analytics (SMA): Social media landscape, Need for SMA; SMA in Small organizations; SMA in large organizations; Application of SMA in different areas.	
Unit II:	**(6 Hrs.)**
Network fundamentals and models: The social networks perspective - nodes, ties and influencers, Social network and web data and methods. Graphs and Matrices- Basic measures for individuals and networks. Information visualization	
Unit III:	(6 Hrs.)
Making connections: Link analysis. Random graphs and network evolution. Socialcontexts: Affiliation and identity.	
Unit IV:	7 Hrs.)
Web analytics tools: Clickstream analysis, A/B testing, online surveys, Web crawling and Indexing. Natural Language ProcessingTechniques for Micro-text Analysis	
Unit V:	7 Hrs.)
Facebook Analytics: Introduction, parameters, demographics. Analyzing page audience. Reach and Engagement analysis. Post-performance on FB. Social campaigns. Measuring and Analyzing social campaigns, defining goals and evaluating outcomes, Network Analysis. (LinkedIn,Instagram, YouTube Twitter etc. Google analytics. Introduction. (Websites)	
Unit VI:	(7 Hrs.)
Processing and Visualizing Data, Influence Maximization, Link Prediction, Collective Classification, Applications in Advertising and Game Analytics Introduction to Python Programming, Collecting and analyzing social media data; visualization and exploration	
Total Lecture	**39 Hours**

Textbooks:	
1.	Matthew Ganis, Avinash Kohirkar, "Social Media Analytics: Techniques and Insights for Extracting Business Value Out of Social Media"Pearson 2016
2.	Jim Sterne," Social Media Metrics: How to Measure and Optimize Your Marketing Investment "Wiley Latest edition
3.	Oliver Blanchard," Social Media ROI: Managing and Measuring Social Media Efforts in Your Organization (Que Biz-Tech)"Que Publishing Latest edition

Reference Books:				
1.	Marshall Sponder	Social Media Analytics	McGraw Hill	Latest edition
2.	Tracy L. Tuten, Michael R. Solomon	Social Media Marketing	Sage	Latest edition

MOOCs Links and additional reading, learning, video material	
1.	

II Semester
25DS222 : PE-IV : Information Retrieval and Recommendation

Course Outcomes :

Upon successful completion of the course the students will be able to

1. Understand the working, significance, applications of Information retrieval systems.
2. Compare different IR models.
3. Design text and multimedia indexing structures for searching of web documents.
4. Justify the evaluation techniques to measure the performance of Information Retrieval System.
5. Apply machine learning algorithms for information retrieval.
6. Design image retrieval algorithms.

Unit I:	(7 Hrs.)
Introduction: Information Retrieval systems, Working with electronic text, Test Collections, Open source IR systems, Information versus Data Retrieval, Basic Concepts: The Retrieval Process, Logical View of Documents. Modelling: A Taxonomy of IR Models, Reference Collections. Significance of Information Retrieval, Impact of the web on Data Retrieval, Applications of Data Retrieval, Basic Data Retrieval System Architecture, Relationships between Digital library and IRS, Open Source IR Systems : Lucene , Wumpus	
Unit II:	(7 Hrs.)
Basic Searching and Indexing: Preprocessing: Simple Tokenizing, Stop-word Removal, Stemming and Lemmatization, Boolean and vector-space retrieval models, Sparse Vectors, Positional Postings, Inverted (static and dynamic) indices, Index Construction, Index Compression, Term weighting, TF-IDF weighting, cosine similarity, Relevance feedback and query expansion. Language Model based IR, Probabilistic Model, Binary Independence Model, Latent Semantic Indexing Mode	
Unit III:	(7 Hrs.)
Evaluation: Data Retrieval System Evaluation, Standard test Collections, Evaluation of Unranked Retrieval Sets, Evaluation of Ranked Retrieval Results, Assessing Relevance, Evaluations on Benchmark Text Collections. The Text Retrieval Conference (TREC), Using Statistics in Evaluation, Minimizing Adjudication Effort, Non-traditional Effectiveness Measures, Measuring Efficiency: Efficiency Criteria, Queueing Theory, Query Scheduling, Caching	
Unit IV:	(6 Hrs.)
Web Search: Web Search Basics, Web Crawling and Indexing, XML retrieval, Link Analysis, Page Rank and HITS algorithms, Searching and Ranking, Relevance Scoring and ranking for Web, Hubs and Authorities. Multimedia IR: Spatial Access Methods, Distance Function, Generic Multimedia Indexing Approach	
Unit V:	(6 Hrs.)
Parallel and distributed IR: Hadoop and Map Reduce, Personalized search, Collaborative filtering and content-based recommendation of documents and products, handling "invisible" Web, Snippet generation, Summarization, Question Answering, Cross-Lingual Retrieval. Vector space classification, Support vector machines and machine learning on documents, Flat clustering, Hierarchical clustering, Matrix decomposition. Naive Bayes, Decision Trees, and Nearest Neighbor, expectation maximization (EM).	
Unit VI:	(6 Hrs.)
Image Retrieval: Content-based Image Retrieval, Image Feature Description, Order system, Texture, Shape, Characteristics of Image Queries, Image Retrieval systems.	

Total Lecture	**39 Hours**

Textbooks:	
1.	C. Manning, P. Raghavan, and H. Sch¨utze, Introduction to Information Retrieval, Cambridge University Press, 2008.
2.	Ricardo Baeza -Yates and Berthier Ribeiro – Neto, Modern Information Retrieval: The Concepts and Technology behind Search 2nd Edition, ACM Press Books 2011
3.	Stefan B¨uttcher, Charles L, A. Clarke, Gordon V. Cormack, Information Retrieval: Implementing and evaluating search engines, MIT Press,2010
4.	Information Storage and Retrieval Systems: Theory and Implementation by Gerald J. Kowalski, Mark T. Maybury , Second Edition, Kluwer Academic Publishers.

Reference Books:	
1.	David A. Grossman, Ophir Frieder, Information Retrieval: Algorithms and Heuristics, Springer, 2004
2.	Frakes, Information Retrieval: Data Structures and Algorithms, Pearson, 2009
3.	Bruce Croft, Donald Metzler and Trevor Strohman, Search Engines: Information Retrieval in Practice, 1st Edition Addison Wesley, 2009.
4.	Mark Levene, An Introduction to Search Engines and Web Navigation, 2nd Edition Wiley, 2010.
5.	Modern Information Retrival By Yates Pearson Education.

MOOCs Links and additional reading, learning, video material	
1.	rijsbergen79_infor_retriev.pdf
2.	https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

II Semester
25DS224 : PE IV: Social Media Analytics

Course Outcomes :
Upon successful completion of the course the students will be able to 1. Familiarize the learners with the concept of social media analytics and understand its significance. 2. Familiarize the learners with the tools of social media analytics. 3. Enable the learners to develop skills required for analyzing the effectiveness of social media for business purposes

Unit I:	(6 Hrs.)
Introduction to Social Media Analytics (SMA): Social media landscape, Need for SMA; SMA in Small organizations; SMA in large organizations; Application of SMA in different areas.	
Unit II:	(7 Hrs.)
Network fundamentals and models: The social networks perspective - nodes, ties and influencers, Social network and web data and methods. Graphs and Matrices- Basic measures for individuals and networks. Information visualization	
Unit III:	(7 Hrs.)
Making connections: Link analysis. Random graphs and network evolution. Socialcontexts: Affiliation and identity.	
Unit IV:	(6 Hrs.)
Web analytics tools: Clickstream analysis, A/B testing, online surveys, Web crawling and Indexing. Natural Language ProcessingTechniques for Micro-text Analysis	
Unit V:	(7 Hrs.)
Facebook Analytics: Introduction, parameters, demographics. Analyzing page audience. Reach and Engagement analysis. Post- performance on FB. Social campaigns. Measuring andAnalyzing social campaigns, defining goals and evaluating outcomes, Network Analysis. (LinkedIn, Instagram, YouTube Twitteretc. Google analytics. Introduction. (Websites)	
Unit VI:	(6 Hrs.)
Processing and Visualizing Data, Influence Maximization,Link Prediction, Collective Classification, Applications in Advertising and Game Analytics Introduction to Python Programming, Collecting and analyzing social media data; visualization and exploration	
Total Lecture	39 Hours

Textbooks:	
1.	Matthew Ganis, Avinash Kohirkar, "Social Media Analytics: Techniquesand Insights for Extracting Business Value Out of Social Media" Pearson, 2016
2.	Jim Sterne," Social Media Metrics:How to Measure and Optimize Your Marketing Investment" Wiley, Latest edition
3.	Oliver Blanchard," Social Media ROI:Managing and Measuring Social Media Efforts in Your Organization (Que Biz-Tech)" Que Publishing, Latest edition

Reference Books:	
1.	Marshall Sponder, Social Media Analytics, McGraw Hill, Latest edition
2.	Tracy L. Tuten, Michael R. Solomon, Social Media Marketing, Sage, Latest edition

Website / Data sheet:

SN	Title
1	Indian Journal of Marketing
2	The Journal of Social Media in Society
3	Social Networks
4	Journal of Digital and Social Media Marketing
5	Social Media Marketing (Magazine)
6	Brand Equity – Economic Times

III SEMESTER
25DS301 : Project Phase – I

IV SEMESTER
25DS401 : Project Phase - II