



**PSGR
Krishnammal College for Women**



**College of Excellence,  2023 - 4th Rank
Autonomous and Affiliated to Bharathiar University
Reaccredited with A++ grade by NAAC, An ISO 9001: 2015 Certified Institution
Peelamedu, Coimbatore-641004**

DEPARTMENT OF DATA ANALYTICS (PG)

**CHOICE BASED CREDIT SYSTEM (CBCS) &
LEARNING OUTCOME BASED CURRICULAR FRAMEWORK (LOCF)**

**MASTER OF DATA ANALYTICS
2023 – 2025 BATCH ONWARDS**



PROGRAMME OUTCOMES

PO1: Demonstrate understanding of the core concepts of analytics to discover data-driven insights

PO2: Illustrate higher degree of technical skills that enable business decision making

PO3: Apply analytical and managerial skills to enhance employment potential

PO4: Exhibit holistic development with emphasis on values and ethics

PROGRAMME SPECIFIC OUTCOME

Graduates will be able to

PSO1: Demonstrate understanding of concepts in data science, statistical concepts and probability.

PSO2: Identify and analyze complex issues reaching substantiated conclusions using the techniques in data science.

PSO3: Design and propose innovative solutions for complex problems that meet the specified business needs.

PSO4: Ability to understand the industry requirements and to have sound knowledge about the professional skills required for data science.

PSO5: Create, select and apply appropriate techniques, tools, resources in data science for prediction and modeling of complex activities with an understanding of the limitation.

PSO6: Communicate effectively on complex tasks in profession as well as with society at large, such as, being able to comprehend and write effective reports, make effective presentations and provide as well as receive clear instructions.

PSO7: Apply ethical principles in research and commit to professional ethics and responsibilities.

PSO8: Recognize the need for lifelong learning and have the ability to engage in independent learning keeping in mind the rapid technological changes.



College of Excellence, **nirf** 2023 - 4th Rank
Autonomous and Affiliated to Bharathiar University
Reaccredited with A++ grade by NAAC, An ISO 9001: 2015 Certified Institution
Peelamedu, Coimbatore-641004

DEPARTMENT OF DATA ANALYTICS (PG)
CHOICE BASED CREDIT SYSTEM & LEARNING OUTCOME BASED
CURRICULAR FRAMEWORK SYLLABUS&SCHEME OF EXAMINATION

MASTER OF DATA ANALYTICS – 2023-2025 BATCH ONWARDS
SEMESTER I & II

Semester	Part	Subject Code	Title of paper	Instruction Hrs / week	Contact Hrs	Tutorial Hrs	Duration of Examination	Examination Marks			Credits
								CA	ESE	Total	
I	III	MDA2301	Paper 1: Descriptive Statistics	4	58	2	3	25	75	100	4
I	III	MDA2302	Paper 2: Foundations of Data Science	4	58	2	3	25	75	100	4
I	III	MDA2303	Paper 3: Linear Algebra	4	58	2	3	25	75	100	4
I	III	MDA2304	Paper 4: Data Structures	4	58	2	3	25	75	100	4
I	III	MDA2305	Paper 5: Information Retrieval	4	58	2	3	25	75	100	4
I	III	MDA23P1	Lab1: RDBMS Lab	5	75	-	3	25	75	100	3
I	III	MDA23P2	Lab2: Full Stack development Lab	5	75	-	3	25	75	100	3
I	III		Online Course	-	-	-	-	-	-	-	-
II	III	MDA2306	Paper 6: Probability and Distributions	5	73	2	3	25	75	100	5
II	III	MDA2307	Paper 7: Machine Learning	4	58	2	3	25	75	100	4
II	III	MDA2308	Paper 8: Big Data Framework	4	58	2	3	25	75	100	4
II / III	III	MDA23CE*	Coursera: Predictive Analytics /	3	45	-	-	100	-	100	3
		MDA2313#	Special Course: Business Analytics	3	43	2	2	100	-	100#	3

II	III	MDA23E2 / MDA23E3	Elective – I: Marketing Analytics Elective – I: Text Analytics	4	58	2	3	25	75	100	4
II	III	MDA23P3	Lab 3: Data Analytics Lab II(Hadoop, MapReduce & R,SPSS)	5	75	-	3	25	75	100	3
II	III	MDA23P4	Lab 4: Machine learning Lab	5	75	-	3	25	75	100	3
II	III		Online Course	-	-	-	-	-	-	-	-

*** Coursera Course**

No ESE, only Continuous assessment of 100 marks will be conducted

List of Electives:

S. No	Course Code	Course Title
Semester II – Elective 1		
1	MDA23E1	Artificial Intelligence and Analytics in Business
2	MDA23E2	Marketing Analytics
3	MDA23E3	Text Analytics
Semester III – Elective 2		
4	MDA23E4	IOT Analytics
5	MDA23E5	Social Media Analytics
6	MDA23E6	Web Data Analytics

CA Question from each unit comprising of

One question with a weightage of 2 Marks :2 x 3 = 6

One question with a weightage of 5 Marks (Internal Choice at the same CLO level) :5 x 3 =15

One question with a weightage of 8 Marks (Internal Choice at the same CLO level) :8 x 3 =24

Total :45 Marks

ESE Question Paper Pattern: 5 x 15 = 75 Marks

Question from each unit comprising of

One question with a weightage of 2 Marks: 2 x 5=10

One question with a weightage of 5 Marks (Internal Choice at the same CLO level): 5 x 5 =25

One question with a weightage of 8 Marks (Internal Choice at the same CLO level): 8 x 5 =40

Continuous Internal Assessment Pattern

Theory

I Year PG (23 Batch)

- CIA Test: 5 marks (conducted for 45 marks after 50 days)
 - Model Exam: 7 marks (Conducted for 75 marks after 85 days (Each Unit 15 Marks))
 - Seminar/Assignment/Quiz: 5 marks
 - Class Participation: 5 marks
 - Attendance: 3 marks
- Total: 25 Marks**

Practical Lab

- Performance: 7 marks
 - Regularity: 5 marks
 - Model Exam: 10 marks
 - Attendance: 3 marks
- Total: 25 marks**

ESE Practical Pattern

The End Semester Examination will be conducted for a maximum of 75 marks respectively with a maximum 15 marks for the record and other submissions if any.

Project:

Evaluation of Individual / Group Project & Viva Voce

I Review - Selection of the field of study, Topic & literature collection	: 5 Marks
II Review - Research Design &Data Collection	: 10 Marks
III Review - Analysis Conclusion Preparation of rough draft	: 10 Marks
Total : 25 Marks	

End semester examination:

Evaluation of the project: 25 Marks
Viva Voce: 50 Marks
Total: 75 Marks

MDA2301	DESCRIPTIVE STATISTICS	Category	L	T	P	Credit
		Theory	58	2	-	4

Preamble

The course introduces the measures of central tendency and dispersion. It also provides the students with systematic knowledge in correlation, regression and outlier analysis in R Programming.

Course Learning Outcomes

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

CLO Number	CLO Statement	Knowledge Level
CLO1	Demonstrate the basic concepts of statistics	K2
CLO2	Apply the concepts of correlation, regression	K3
CLO3	Identify the methods for different measures of central tendency, dispersion	K4
CLO4	Evaluate the methods for representation of data.	K5
CLO5	Construct various plots, outliers for regression, diagnostics	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	M	S	S
CLO2	S	S	S	M
CLO3	S	S	M	S
CLO4	M	S	S	S
CLO5	S	S	M	S

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(12 Hrs)

Introduction: **Nature and scope of Statistics, limitations of statistics** - Types of data: Concept of population and sample, primary and secondary data, quantitative and qualitative data, discrete and continuous data, cross-sectional and time series data. **Scales of measurement: Nominal, Ordinal, Ratio and Interval.** Case Studies using Statistical Methods in Excel.

UNIT II

(12 Hrs)

Tabulation of Data- Difference between Classification and Tabulation – Parts of Tabulation –Rules for Tabulation – Types of Tables

Diagrammatic representations: - Line diagram, bar diagram, pie diagram and sub-divided bar diagram, Frequency distribution and cumulative frequency distribution and their graphical

representations, Frequency polygon, histogram, ogive, frequency curves, stem and leaf displays. Exploring R Basics – R Features – R Basic Data types – R Basic Operators – Data visualization in R

UNIT–III (12 Hrs)

Univariate data: Different measures of location, dispersion, relative dispersion, **skewness and kurtosis**, Moments, Quantiles and measures based on them – comparison with moment Measures - **Box-plot and detection of outliers. Trimmed mean and Winsorised mean** – Simple problems. Interpretation of Histogram, Box Plot in R - Outliers on Univariate Data in R

UNIT IV (11 Hrs)

Correlation – Scatter diagram – Karl Pearson Coefficient of Correlation -Spearman's Rank coefficient of Correlation - **Rank Correlation Coefficient**. Interpretation of Scatter Plot in R

UNIT V (11 Hrs)

Regression: Introduction – Uses of regression analysis – **regression lines – regression equations of X on Y and Y on X – regression equation in terms of correlation table**. Visualization plots for Linear Regression in R – Interpretation of Logistic Regression in R.

Text Book

1. S.C.Gupta ,V.K.Kapoor (2019), Fundamentals of Mathematical Statistics, Sultan Chand & Sons (Unit I : Chapter 1 , Price Michael (2012) , Excel 2010 In Easy Steps, Tata MCGraw - Hill - Chapter 8 (123 -133,135)
2. R.S.N. Pillai and Bhagavathi (2019) , Statistics Theory and Practice, S. Chand Publishing Company Pvt Ltd (Unit II : Chapter 6: Pg 54 – 73, Chapter 7, Chapter 8: 100-112) (Unit IV :Chapter 12: Pg 396 - 420)
3. R.Wilcox (2009), Basic Statistics, Oxford University Press, (Unit III: 2.1-2.5)
4. S.P.Gupta (2021), Statistical Methods, Sultan Chand and Sons (Unit V: Chapter 11)
5. V.Bhuvanewari(2018), Data Analytics with R Programming, Scitech Publications Pvt Ltd (Unit II :3-3.2,3.7-3.8, 4.3 Unit III : 4.4, 4.6, 9.2.1, Unit IV: 4.7,4.11, Unit V:6.3,6.9)

Reference Books

1. Murray R Spiegel and Larry J Stephens: Statistics, Schaum's Outline, Fourth edition, 2008
2. R.S.N. Pillai, Statistics, S. Chand Publishing Company Pvt Ltd , 1992
3. <https://www.indiabix.com/data-interpretation/questions-and-answers/>
4. <https://www.mathsisfun.com/data/pictographs.html>

Pedagogy

Lectures, Simulation exercises, Demonstration

Course Designers

1. Dr. T. A. Albinaa
2. Dr. N. Radha

MDA2302	FOUNDATIONS OF DATA SCIENCE	Category	L	T	P	Credit
		Theory	58	2	-	4

Preamble

This course introduces the essential knowledge about foundations of data science with python. It provides various steps involved in the data science process and core python with advanced concepts like regular expressions, exception handling, multithreading. It also introduces the python libraries that are most commonly used for data analysis.

Prerequisite

- Basic understanding of Open source software
- Database concepts

Course Learning Outcomes

On successful completion of the course, the students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Understand data classification, process of big data technology, user roles and skills in data science and its tools.	K2
CLO2	Apply the fundamental concepts, tools and techniques of data science in 360 view of Customer.	K3
CLO3	Analyze the methodologies of data science and its tools.	K4
CLO4	Evaluate the utilities and functionalities of Python	K5
CLO5	Design solutions for data analytics problems using Python.	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	M	M	M
CLO2	S	S	M	M
CLO3	S	S	S	S
CLO4	S	S	S	S
CLO5	S	S	S	S

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(12 Hrs)

Data Evolution: Data Growth- a Perspective - IT Components - Business Process – Landscape - Data to Data Science — Understanding data: Introduction - Types of Data: Numeric - Categorical – Graphical - High Dimensional Data - Data Classification - **Sources of Data: Time Series - Transactional Data - Biological Data -Spatial Data - Social Network Data** - Introduction to Python programming – Libraries – Plugins – Problem solving.

UNIT II

(12 Hrs)

Introduction to Data Science - **A Discipline -Data Science vs Statistics, Data Science vs Mathematics**, Data Science vs Programming Language, Data Science vs Database, Data Science vs Machine Learning. Data Analytics - Data Manipulation with NumPy - Introduction to NumPy arrays and their operations-

Manipulating arrays: indexing, slicing, and reshaping - Working with NumPy, Working with statistical toolbox.

UNIT III

(11 Hrs)

Relation: Data Science, Analytics and Big Data Analytics. Data Science Components –Big data technology - Data Science user - roles and skills - **Data Science use cases.** Introduction to the Pandas library - Loading and exploring datasets - **Data indexing and selection** – Data cleaning and preprocessing - Aggregation and summarization - Data ingestion: export/import data from various file formats (CSV, Excel, etc.).

UNIT IV

(11 Hrs)

Digital Data-an Imprint: Evolution of Big Data –What is Big Data –Sources of Big Data. **Characteristics of Big Data 6Vs –Big Data Myths** –Data Discovery-Traditional Approach. Introduction to data visualization - Plotting with Matplotlib: line plots, scatter plots, bar plots, and histograms – **Exploratory data analysis** - Data visualizations to explore and analyze datasets.

UNIT V

(12 Hrs)

Big data Technology: Big Data Technology Process –**Big Data Exploration -Data Augmentation – Operational Analysis –Security and Intelligence.** Introduction to NLTK-Installing NLTK and downloading necessary Resources-Basic NLTK functionalities and data structures-Tokenization and Text Preprocessing: **Text cleaning and normalization techniques** - Exploring NLTK for text processing applications like sentiment analysis.

Text Book

1. V. Bhuvanewari, T. Devi, (2016). Big Data Analytics: A Practitioner’s Approach, Bharathiar University
2. Wesley J.Chun (2016). Core Python Programming Application Programming, 3/e, Pearson Education Publication, United States

Reference Books

1. Mark Lutz (2010). Programming Python, 4/e, O’Reilly Media.
2. Wes McKinney (2022), Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, Grayscale Indian Edition.
3. Alberto Boschetti, Luca Massaron (2016), Python Data Science Essentials, Packt Publishing.
4. Wesley J.Chun (2006). Core Python Programming, 2/e, Prentice Hall Publication.

Pedagogy: Lectures, Demonstration and Case Studies

Course Designers

1. Mrs. K. Gandhimathi
2. Mrs. G. Anitha

MDA2303	LINEAR ALGEBRA	Category	L	T	P	Credit
		Theory	58	2	-	4

Preamble

The course introduces the principles underlying linear equations and vector spaces. It also provides the concepts of Eigen values, Eigen vectors and Positive Definite Matrices in Scilab.

Course Learning Outcomes

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

CLO Number	CLO Statement	Knowledge Level
CLO1	Demonstrate competence with the basic ideas of linear Algebra including the concepts of vector spaces, Determinants, Eigen values and Eigenvectors and positive definite matrices	K2
CLO2	The ability to understand the principles of Linear Algebra	K3
CLO3	Apply properties of linear spaces to specific mathematical structures	K4
CLO4	Compose clear and accurate proofs using the concepts of linear Algebra	K5
CLO5	Appreciate the significance of vector spaces and positive definite matrices	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	S	S	S
CLO2	S	S	S	M
CLO3	S	S	S	S
CLO4	S	S	M	S
CLO5	S	S	S	S

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(12 Hrs)

The Geometry of Linear Equations- An Example of Gaussian Elimination- Matrix Notation and Matrix Multiplication - **Triangular Factors and Row Exchanges**- Inverses and Transposes – Implementation of basic matrix operations(Sum, Multiplication, Square root, Cube, Inverse, Transpose, Rank, Fibonacci) in scilab

UNIT II

(11 Hrs)

Vector Spaces: **Vector Spaces and Subspaces** – Solving $Ax=0$ and $Ax=b$ - Linear Independence, Basis, and Dimension- The Four Fundamental Subspaces- **Graphs and Networks**- Linear Transformations, Implementation of vector space and subspace, Linear independency, to find the reduced row echelon form of a matrix in Scilab.

UNIT III

(11 Hrs)

Determinants: Introduction- **Properties of the Determinant**- Formulas for the Determinant- Applications

of Determinants – Implementation of determinants, Properties - associative, commutative and distributive property in a matrix using Scilab

UNIT IV

(12 Hrs)

Eigenvalues and Eigenvectors: Introduction- Diagonalization of a Matrix. - Difference Equations and Powers A^k - Differential Equations and e^{At} - **Complex Matrices- Similarity Transformations** – Implementation of Eigenvalues and eigenvectors, solving equations by Gauss elimination, Gauss Jordan Method and Gauss Siedel in Scilab

UNIT V

(12 Hrs)

Positive Definite Matrices: Minima, Maxima, and Saddle Points - Tests for Positive Definiteness- **Singular Value Decomposition, Minimum Principles**, The Finite Element Method, Implementation of Minima, Maxima, Saddle Points and plot the function of the derivatives in Scilab

Text Book

Gilbert Strang(2020). Linear Algebra and Its Application, Fourth Edition, Academic Press.

Reference Books

1. David C. Lay, Steven R. Lay, Judi J. McDonald (2016). Linear Algebra and Its Applications, Pearson Education.
2. Peter D. Lax(2014). Linear Algebra and Its Applications, Second Edition, Wiley Publication

Pedagogy: Lectures, Demonstration and Case Studies

Course Designers

1. Dr. T.A.Albinaa
2. Dr. T.Brindha

MDA2304	DATA STRUCTURES	Category	L	T	P	Credit
		Theory	58	2	-	4

Preamble

This course covers the various data structures, including arrays, structures, stacks and queues. It includes sorting and searching techniques and effective search methods in Binary trees. This course also deals with graph data structures.

Prerequisite

- Discrete mathematics.

Course Learning Outcomes

On successful completion of the course, the students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Understand the concepts of arrays, strings and algorithms for basic operations.	K2
CLO2	Apply concept of stacks, queues, linked list and algorithms for basic operations.	K3
CLO3	Identify the familiarity with major algorithms and data structures	K4
CLO4	Analyze appropriate algorithms and data structures for various applications	K5
CLO5	Formulate the computational complexity of various algorithms	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	M	M	M
CLO2	S	M	M	M
CLO3	S	S	M	M
CLO4	S	S	S	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(12 Hrs)

Basics: Algorithm Specification – Data Abstraction – Performance Analysis:Time, Space Complexity – Arrays – Dynamic Allocated Arrays – Structures and Unions – Polynomials - **Sparse Matrices- Representation of Multidimensional Arrays – Strings.**

UNIT II

(12 Hrs)

Stacks and Queues: Stacks – Stacks Using Dynamic Arrays - Queues - Circular Queues Using Dynamic Arrays - Evaluation of Expressions - **Multiple Stacks And Queues - Recursion.**
Linked Lists: Singly Linked List And Chains – Representing Chains in C – Linked Stack And Queues – **Polynomials - Additional List Operations - Sparse Matrices – Doubly Linked List.**

UNIT III

(11 Hrs)

Searching: Introduction - Sequential Search - **Binary Search** – **Analysis.Sorting:**Introduction - Insertion Sort - Selection Sort - Merge Sort - **Quick Sort - Heaps and Heap Sort - Analysis.**

UNIT IV

(11 Hrs)

Efficient Binary Search Trees:Binary Tree – Traversals - Optimal Binary Search Trees – AVL Trees- **KD Trees.**

UNIT V

(12 Hrs)

Graphs: The graph Abstract Data Type- Elementary graph operations- Minimum cost spanning trees- shortest paths and transitive closure- **AOV networks –AOE networks.**

Case Study: Location Identification, Game Development, Google Knowledge Graph, Path Optimization

Text Book

1. Ellis Horowitz, Sartaj Sahni and Anderson Freed (2009), Fundamentals of data structures in C, University Press
2. Ellis Horowitz, Sartaj Shani, SanguthevarRajasekaran (2013), Fundamentals of computer algorithms, Galgotias Publications private limited
3. Robert L Kruse (2008). Data Structures & Program Design, New Delhi: Prentice Hall

Reference Books

1. Mark Allen Weiss(2012). Data Structures and Algorithm Analysis in C++; Pearson Education
2. Adam Drozdek (2013). Data Structures and Algorithm in C++ , 4th Edition.
3. Ch. Rajaramesh C.V. Sastry, Rakesh Nayak (2019), Data Structures and Algorithms, Wiley.
4. Akhilesh Kumar Srivastava (2019), A Practical Approach to Data Structure and Algorithm with Programming in C.

Pedagogy: Lectures, Group Discussion, case study

Course Designers

1. Dr. N. Radha
2. Mrs. G. Anitha

MDA2305	INFORMATION RETRIEVAL	Category	L	T	P	Credit
		Theory	58	2	-	4

Preamble

This course presents the concepts of document representation, document indexing, digital information storage, retrieval and distribution. It also introduces effective search strategies for IR systems, vector space model, text classification and evaluation methods of IR systems.

Prerequisite

- Statistics
- Basic Programming

Course Learning Outcomes

On the successful completion of the course, students will be able to

CO Number	CLO Statement	Knowledge Level
CLO1	Demonstrate the concepts of document representation, document indexing, digital information storage, retrieval and distribution	K2
CLO2	Apply the concepts of vector spaces and classifiers to perform document classification.	K3
CLO3	Examine the advantages and disadvantages of different information-retrieval models	K4
CLO4	Determine the effective search strategies for IR systems	K5
CLO5	Able to develop information retrieval system for specific use cases.	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1.	S	M	M	M
CLO2.	S	M	M	M
CLO3.	S	S	M	M
CLO4.	S	M	S	M
CLO5.	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(12 Hrs)

Boolean retrieval: Information retrieval problem - Processing Boolean queries - Boolean model versus ranked retrieval. **The term vocabulary** and postings lists: Document delineation and character sequence decoding - Determining the vocabulary of terms - Faster postings list intersection via **skip pointers** - **Positional postings and phrase queries**

UNIT II

(12 Hrs)

Dictionaries and tolerant retrieval: Search structures for dictionaries - **Wildcard queries** - Spelling correction - Phonetic correction. Index construction: Hardware basics - **Blocked sort-based indexing** - **Single-pass in-memory indexing** - Distributed indexing - Dynamic indexing - Other types of indexes

UNIT III

(12 Hrs)

Scoring, term weighting and the **vector space model**: Parametric and zone indexes - Term frequency and weighting - The vector space model for scoring. **Evaluation in information retrieval**: Information Retrieval system – Standard test Collection – Evaluation of unranked retrieval sets – **Evaluation of ranked retrieval results** – Assessing relevance

UNIT IV

(11 Hrs)

XML retrieval: Basic XML concepts – A vector space model for XML retrieval – Evaluation of XML retrieval - Text-centric vs. data-centric XML retrieval. **Text classification and Naive Bayes**: The text classification problem - Naive Bayes text classification - Properties of Naive Bayes - Feature selection - **Evaluation of text classification**

UNIT V

(11 Hrs)

Vector space classification: Document representations and measures of relatedness in vector spaces – **Rocchio classification** - Flat clustering: Clustering in information retrieval - **Evaluation of clustering - K-means** – Web search basics - Web characteristics - Advertising as the economic model – Search user experience – Basic Page Rank

Text Book

Christopher D. Manning, Prabhakar Raghavan, Henrich Schutze (2008). Introduction to Information Retrieval, 1/e; New York: Cambridge University Press

Reference Books

1. Stefan Buttcher et.al (2016). Information Retrieval - Implementing and Evaluating, MIT Press
2. Dr Ricardo Baeza-Yates et.al (2011). Modern Information Retrieval: The Concepts and Technology, Addison Wesley
3. David A. Grossman and Ophir Frieder (2010). Information Retrieval, 2/e, Universities Press

Pedagogy: Lectures, Demonstrations, Guest Lecture, Video Lectures

Course Designers

1. Dr. N.Radha
2. Dr. V. Bharathi

MDA23P1	RDBMS LAB	CATEGORY	L	T	P	CREDITS
		III	-	-	75	3

Preamble

This course provides sound introduction to implement the relational database management systems concepts in SQL. This course also provides various exercises to implement the schema and Table creation, ER diagram, integrity constraints, Joins, Nested queries, functions, procedures, cursors, triggers, exception handling and to create Simple real time applications.

Prerequisite

- Database concepts
- Programming concepts

Course Outcomes

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1.	Implement the databases concepts and SQL queries as per user requirement	K2
CO2.	Apply specific SQL commands on relational tables for different situations	K3
CO3.	Analyse use cases and create constraints suitable for the given situation.	K4
CO4.	Create and analyse a database using SQL DML/DDDL commands	K6
CO5.	Design and build a GUI application	K6

Mapping with Programme Outcomes

Cos	PO1	PO2	PO3	PO4
CO1.	S	S	S	S
CO2.	S	S	S	S
CO3.	S	S	S	S
CO4.	S	S	S	S
CO5.	S	S	S	S

S- Strong; M-Medium; L-Low

Syllabus

- Exercises to create schema, Table, ER Diagram using online tool
- Exercises to implement the concepts of null constraint, unique constraint, integrity constraints, check constraints, default constraints.
- Exercises to implement nested queries.
- Exercises to implement the concepts of partitioning queries
- Exercises to create a view from the tables
- Exercises to create functions and procedures
- Exercise to create triggers and queries
- Exercises to create cursors and exceptions

- List of Projects using frontend and backend (Topics not limited to)
 1. Student Admission Management
 2. Library Management System
 3. Payroll Management System
 4. Telecom
 5. Online Retail Application Database
 6. Inventory Control Management
 7. Hospital Management System
 8. Banking system management
 9. Railway ticket reservation management system
 10. An SMS-based Remote Server Monitoring System
 11. Blood Donation Database
 12. Cooking Recipe Portal
 13. Electricity Bill payment system
 14. Online tax management system
 15. Training and placement system

Course Designers:

1. Dr.N.Radha
2. Dr.S. Meera

MDA23P2	FULL STACK DEVELOPMENT LAB	CATEGORY	L	T	P	CREDITS
		III	-	-	75	3

Preamble

This lab course is intended to explore concepts into full stack development through HTML, CSS, javascript and Node.js. It makes the students to learn various designing approaches and techniques to provide a robust UI/UX experience using various software tools to implement web pages, style sheets, forms, events, server side applications.

Course Learning Outcomes

On the successful completion of the course, students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Understand the web design elements, functions, files, data connections, forms, events.	K2
CLO2	Apply web design methods to solve problems	K3
CLO3	Analyze the web page design requirements and design web pages.	K4
CLO4	Create a web application using HTML, CSS, javascript and Node.js.	K6
CLO5	Design and build web application using HTML, CSS, Javascript and Node.js	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	S	S	S
CLO2	S	S	S	S
CLO3	S	S	S	S
CLO4	S	S	S	S
CLO5	S	S	S	S

S- Strong; M-Medium; L-Low

LIST OF PROGRAMS

- Exercise to design a website with forms, frames, links and table using HTML tag.
- Exercise to build a countdown timer using HTML, CSS and JavaScript to display the remaining time until a specific event.
- Exercise to design a navigation menu with HTML unordered lists and CSS styles to create a dropdown or responsive menu.
- Exercise to design responsive web page that adjusts its layout and design based on the screen size using CSS media queries.

- Exercise to design a platform for embedding video from you tube using iframe.
- Exercise to design web server using built-in module in Node.js.
- Exercise to develop a HTML and CSS program to find the average and grade of students marks.
- Exercise to implement an image gallery using a grid layout and CSS to showcase a collection of image.
- Exercise to develop and demonstrate JavaScript with POP-UP boxes and functions.
- Exercise to build a CRUD operation for a product using laravel framework.

List of Projects using web technologies (Topics not limited to):

1. E-Commerce website
2. Social Media Platform
3. Online Learning Platform
4. Task Management Application
5. Weather Application
6. Event Booking Platform
7. Blogging Platform
8. Real-time Chat Application
9. Student feedback system
10. Online Resume Builder

Course Designers:

1. Mrs.K. Gandhimathi
2. Mrs. G. Anitha

MDA2306	PROBABILITY AND DISTRIBUTIONS	Category	L	T	P	Credit
		Theory	73	2	-	5

Preamble

This course introduces the fundamental concepts of probability and random variables. It also provides knowledge in discrete and continuous distributions. It deals with various sampling distributions like t, F, chi-square distributions etc. It also provides the students with systematic knowledge in R Programming.

Course Learning Outcomes

On the successful completion of the course, students will be able to

CLO Number	CO Statement	Knowledge Level
CLO1	Understand the concepts of probability and its distributions.	K2
CLO2	Solve problems involving random variables and their functions	K3
CLO3	Examine the sampling distributions to contribute to the process of making rational decisions in analytical problems.	K4
CLO4	Compare the various sampling tests and choose test suitable for the problem	K5
CLO5	Formulate the hypothesis test for various complex problems	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	M	M	M
CLO2	S	S	M	M
CLO3	S	S	S	M
CLO4	S	M	S	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I (14 Hrs)

Random Experiment: **Sample space, Different types of events**. Approaches of measuring probability: Mathematical, Statistical and Axiomatic probability, Laws of probability-conditional probability – Baye's theorem

Interpretation of different types of events in R -Equally Likely Model in R - Baye's Principle in R.

UNIT II (14 Hrs)

Random variable - **Discrete and Continuous Random variables** - Distribution function and its properties – Expectation and Moment Generating function: Raw and Central moments-relationship between central and raw moments-moments about an arbitrary value-moment generating function-**properties of moment generating function**-characteristic function.

Interpretation of Mean and Variance of a Random variable in R -Distribution function in R – Exercises in R.

UNIT III**(15 Hrs)**

Discrete Distributions: Binomial, Poisson, Geometric, Uniform distributions and their properties.
Continuous Distributions: Rectangular, Exponential, Normal distributions and their properties-
Simple problems

Interpretation of Binomial Distributions in R - Poisson Distributions in R - Continuous Distributions in R.

UNIT IV**(15 Hrs)**

Central limit theorem, Liapounoff Form, Lindberg Levy Form. Sampling Distributions –Distribution of Sample Mean, t-Distribution, Chi-square Distribution, Distribution of Sample Variance, F-Distribution, Degrees of freedom, Interval Estimation, Confidence interval for one mean and difference of two means.
Testing of Hypothesis : Relation between confidence interval and testing of hypothesis-level of significance and p-value. Classification of hypothesis tests.

Interpretation of t-Distribution in R - Chi-square Distribution in R - F Distribution in R.

UNIT V**(15 Hrs)**

Large sample tests: Single mean, Difference of two means, Single proportion. **Small sample tests: t-test for single mean, difference between two means-F-test for equality of two population variances-Chi-square test for single Variance**, Chi-square test for goodness of fit- Chi-square test for independence of attributes and homogeneity and equality of proportions.

Interpretation of Large Sample and Small sample test in R. Simple Problems in R

Text Books

Ravichandran. J (2019). Probability and Statistics for Engineers, Wiley.

Unit I: Section: 1.1-1.4

Unit II: Section: 2.1-2.5, 3.1-3.2, 3.3.1 - 3.3.6

Unit III: Section: 4.1,4.2.1-4.2.4,5.1-5.3,5.5

Unit IV: Section: 6.4, 10.1,10.2.1,10.2.2,10.2.3,10.2.4, 10.3.1, 10.3.2, 11.1-11.3

Unit V : Section – 11.4.1-11.4.3, 11.5.1-11.5.7

V.Bhuvanewari(2018), Data Analytics with R Programming, Scitech Publications Pvt Ltd

Reference Books

1. Rohatgi V.K(2015). An Introduction to Probability Theory and Mathematical Statistics, John Wiley.
ScymourLipschuts (2011). Probability, Schaum's Outline.
2. S.C.Gupta and V.K.Kapoor (2018). Fundamentals of Mathematical Statistics, Sultan Chand and Sons.
3. Goon A.M., Gupta M.K. & Dasgupta B (2016). An Outline of Statistical Theory (Vol-1), World Press.
4. V.Bhuvanewari(2018), Data Analytics with R Programming, Scitech Publications Pvt Ltd

Pedagogy : Lectures, Case studies, Demonstrations

Course Designers:

1.Dr.T.A.Albinaa

2.Dr.C.R.Parvathy

MDA2307	MACHINE LEARNING	Category	L	T	P	Credit
		Theory	58	2	-	4

Preamble

This course introduces the fundamentals of Machine Learning and algorithms. It also covers various supervised and unsupervised learning algorithms used for classification, prediction and clustering.

Prerequisite

- Statistics and probability
- Linear Algebra

Course Learning Outcomes

On the successful completion of the course, students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Understand the concepts of machine learning.	K2
CLO2	Apply supervised and unsupervised learning algorithms for classification, prediction and clustering.	K3
CLO3	Analyze the logic behind the execution of various classifiers.	K4
CLO4	Evaluate the performance of different algorithms available for mining data.	K5
CLO5	Predict solution for real world problems.	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	M	M	M
CLO2	S	M	M	M
CLO3	S	S	S	M
CLO4	S	S	S	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(11 Hrs)

Introduction: Machine learning – Examples and Applications - Perspectives and Issues in Machine learning - **Input: Concepts, Instances, and Attributes - Output: Knowledge Representation-Credibility: Evaluating What's Been Learned: Training and Testing - Predicting Performance - Cross Validation - Other Estimates - Counting the cost.**

UNIT II

(11 Hrs)

Decision Tree Learning: Decision tree representation – Decision tree learning –Issues in decision tree learning- Random forest -**Bayesian Learning: Naïve Bayes classifier** - Instance Based Learning:

Introduction – k- nearest neighbor Learning - **Radial Basis Function, Case based reasoning.**

UNIT III

(12 Hrs)

Artificial Neural Network - Introduction – Neural Network Representation - Perceptrons –Gradient Descent and Delta Rule-**Multilayer Networks and Backpropagation Algorithm–Logistic Regression** - Linear models for Regression- Linear Discriminant Analysis.

Use Cases: Spam Filters, Price Prediction, Demand Forecasting, Image Recognition

UNIT IV

(12 Hrs)

SVM : Introduction – Kernel methods - formulation and computation- SVM Linear classifier – SVM with two variables –non-linear classifier-Polynomial kernels- Radial Basis Function Kernels

UNIT V

(12 Hrs)

Clustering Methods.- Introduction – K- Means- Expectation-Maximization Algorithm- Hierarchical Clustering - Choosing the Number of Clusters – **Gaussian Mixture Models – Frequent Pattern (FP) Growth – PCA – Kernal PCA.**

Use cases: Customer Segmentation, Anomaly detection, Recommender System.

Reference Books

1. Ian Witten, Data mining: Practical Machine Learning Tools and Techniques, Fourth edition, Morgan Kaufmann Publishers (2016)
2. Tom M. Mitchell (2013). Machine Learning, Tata McGraw-Hill, New Delhi
3. K.P. Soman, Machine Learning with SVM and Other Kernel Methods, 2011, PHI Publishing
4. Dr.Pablo Rivas, Deep Learning for Beigneers, Packet Publishing, 2020
5. Aston Zhang, ZaheryC.Lipton, Muli and Alexander J Smola(2022), Drive into Deep Learning, Amazon Science
6. Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, (2019), “Machine Learning”, (1/e), Pearson Publishers

Pedagogy : Lectures, Case studies, Demonstrations

Course Designers:

1. Mrs.G.Anitha
2. Dr.N.Radha

MDA2308	BIG DATA FRAMEWORK	Category	L	T	P	Credits
		Theory	58	2	-	4

Preamble

This course introduces big data framework, technologies, storage and Hadoop ecosystem. It also deals with the concepts of MapReduce, Hbase ,Scala and MongoDB.

Prerequisite

- Database Management systems
- Programming concepts

Course Learning Outcomes

On successful completion of the course, students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Understand the characteristics of big data, concepts of Hadoop ecosystem and Scala programming.	K2
CLO2	Apply Mapreduce programming model to process big data.	K3
CLO3	Analyze Hbase and Scala and its uses in big data processing.	K4
CLO4	Evaluate techniques used for handling NoSQL databases MongoDB and Hbase.	K5
CLO5	Design programs for big data applications using Hadoop components.	K6

Mapping with Program Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	S	S	M
CLO2	S	S	S	M
CLO3	S	S	S	M
CLO4	S	S	S	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I (11 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 (11 Hrs)
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 – **YARN**

UNIT III (12 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 - Unit test with MR unit- local test

UNIT IV (12 Hrs)
Hbase: Concepts - Hbase Vs RDBMS - **Creating records- Accessing data – Updating and deleting data – Modifying data-** exporting and importing data.
SCALA: Introduction - Classes and objects- Basic types and operators - **built-in control structures** - functions and closures – **inheritance**

UNIT V (12 Hrs)
NoSQL Databases: Introduction to NoSQL- **MongoDB:** Introduction – Data types – **Creating, Updating and deleting documents -Querying** – Introduction to indexing – Capped collections.
Case Studies: Hadoop at Facebook, Nutch Search Engine.

Text Book

1. Tom White (2015). Hadoop: The Definitive Guide, 4/e, O'Reilly Media
2. Judith Hurwitz (2013). Big Data for Dummies, John Wiley & Sons

Reference Books

1. Bill Franks (2012). Taming the Big Data Tidal wave, John Wiley & Sons
2. Martin Odersky, Lex Spoon, Bill Venners(2019), Programming in Scala, 4/e, Artima Press, California.
3. Shashank Tiwari (2013). Professional NoSQL, John Wiley & Sons
4. Shannon Bradshaw, Eoin Brazil & Krishna Chodorow (2020), MongoDB. The Definitive Guide- Powerful & Scalable Data Storage, 3/e, O'Reilly Media
5. Boris lublinsky, Kevin t. Smith, Alexey, Yakubovich(2015). Professional Hadoop Solutions, Wiley
6. Achintya Kumar,(2019), Big Data Black Book: Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization, DreamTech Press

Pedagogy: Lectures, Group Discussions, Case studies

Course Designers

1. Dr. S. Meera
2. Mrs. G. Anitha

MDA23CE	COURSERA - PREDICTIVE ANALYTICS	Category	L	T	P	Credits
		Theory	45	-	-	3

Course contents

(45 Hrs)

S. NO.	COURSE NAME	LINK
1	Predictive Modelling and Analytics (10 Hrs)	https://www.coursera.org/learn/predictive-modeling-analytics
2	Python Data Products for Predictive Analytics Specialization (40 Hrs)	https://www.coursera.org/specializations/python-data-products-for-predictive-analytics#courses
3	Practical Predictive Analytics: Models and Methods (6 Hrs)	https://www.coursera.org/learn/predictive-analytics

MDA2313	BUSINESS ANALYTICS	Category	L	T	P	Credit
		Theory	43	2	-	3

Preamble

This course provides the fundamentals of domain knowledge in the functional areas of Data Analytics. Various important functional areas such as Health care analytics, Banking and Finance, Telecommunication and Retail Analytics are elaborated in this course. Few Use cases are also covered.

Prerequisite

- Foundations of Data Science
- Big data framework

Course Learning Outcomes

Upon successful completion of this course, students should be able to:

CLO Number	CLO Statement	Knowledge Level
CLO1	Compare various domain areas and their challenges	K2
CLO2	Apply the concepts of analytics to make better decisions	K3
CLO3	Examine use cases for different domains.	K4
CLO4	Evaluate the challenges faced in various domains and choose appropriate analytics solutions in all domains	K5
CLO5	Propose suitable analytics solutions as required by the use cases.	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	M	M	M
CLO2	S	M	M	M
CLO3	S	S	M	M
CLO4	S	M	S	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(8 Hrs)

Healthcare analytics – Introduction - Potential contributions - Challenges of healthcare industry - **current and future state of healthcare analytics – top healthcare analytics adaptations**

UNIT II

(9 Hrs)

Banking and Finance: Systems of Banking – Commercial Banking – New Financial Services: Overview of Analytics in Insurance: Key Insurance Analytics – Challenges – **Health Insurance Analytics, Life Insurance Analytics**- Types of Insurance – Housing Finance.

UNIT III

(9 Hrs)

Telecommunication: Introduction - End-User Needs and Demands- **Telecom Business**

UNIT IV (8 Hrs)
Retail analytics – Understanding the new consumer – Marketing in a consumer- driven era - Managing the brand to drive loyalty

UNIT V (9 Hrs)
Case studies: Walmart, Netflix, Facebook, **Uber, Amazon, Kaggle**

Reference Books

1. Dwight McNeill(2013). A Framework for Applying Analytics in Healthcare: What Can Be Learned from Best Practices in Banking , Retail, Politics and Sports, Pearson Education
2. Gomez Clifford(2011). Banking and Finance Theory Law and practice, PHI Learning
3. Patricia L.Saporito(2014). Applied Insurance Analytics: A Framework for Driving More Value from Data Assets, Technologies and Tools, Pearson Education LTD.
4. Anders Olsson(2005). Understanding Changing Telecommunications, Wiley Publications
5. Jennifer LeClaire, Danielle Dahlstrom, Vivian Braun. Business analytics in Retail for dummies, 2nd IBM Limited edition
6. Alistair Croll (2013) Lean analytics: Use Data to Build a Better Startup faster, O Reilly Publishers
7. Bernard Marr (2016). Big Data in Practice – How 45 successful companies used big data analytics to deliver extraordinary results, Wiley Publications
8. Purba Halady Rao (2013). Business Analytics. An application Focus, PHI Learning private ltd.

Pedagogy : Lectures, Group Discussions, Demonstrations, Case studies

Course Designers

1. Mrs. G. Anitha
2. Dr. S. Meera

EVALUATION PATTERN		
CIA	-	10
Case study Discussion	-	20
Mini-Project	-	20
External Viva Voce	-	50

Total - 100

Question paper pattern for CIA

Bloom's Category	Section	Marks		Total
K3, K4	A – 5 out of 8 x 8 marks	40	300 words	50
K4, K5	B – 1 out of 2 x 10 marks	10	500 words	

MDA23E2	Elective 1: MARKETING ANALYTICS	Category	L	T	P	Credit
		Theory	58	2	-	4

Preamble

This course contains the strategies of marketing analytics, forecasting techniques, Customer Needs and retailing. This will also offer the market research tools involved in retailing.

Prerequisite

- Foundations of Data science

Course Learning Outcomes

On successful completion of the course, the students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Understand the concepts of marketing data and effective use of MicrosoftExcel	K2
CLO2	Apply Forecasting Techniques to improve response rates for marketingcampaigns	K3
CLO3	Analyse Market segmentation based on Cluster Analysis	K4
CLO4	Estimate methods for processing marketing data in prediction and marketsegmentation.	K5
CLO5	Formulate solutions in excel for various use cases of marketing analytics methods	K6

Mapping with Programme Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	M	L	M
CLO2	S	L	L	M
CLO3	S	S	L	M
CLO4	M	M	S	M
CLO5	S	M	M	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I

(11 Hrs)

Marketing Data Summarization - Slicing and Dicing Marketing Data with Pivot Tables - **Using Excel Charts to Summarize Marketing Data** - Using Excel Functions to Summarize Marketing Data.

UNIT II

(12 Hrs)

Forecasting Techniques: Simple Linear Regression and Correlation - Using Multiple Regression to Forecast Sales - Forecasting in the Presence of Special Events - **Modeling Trend and Seasonality** - Ratio to Moving Average Forecasting Method - Winter's Method - **Using Neural Networks to Forecast Sale**

UNIT III**(12 Hrs)**

Customer Needs: Conjoint Analysis - Logistic Regression - Discrete Choice Analysis – **Customer Value - Introduction to Customer value, Benefits**

UNIT IV**(12 Hrs)**

Market segmentation: Cluster Analysis - User-Based Collaborative Filtering - Collaborative Filtering - **Using Classification Trees for Segmentation.**

UNIT V**(11 Hrs)**

Retailing and market research tools : Retailing - Introduction to retailing, Market Basket Analysis and Lift - Marketing Research Tools - **Principal Components Analysis**

Text Book

1. Wayne.L.Winston (2014). Marketing Analytics: Data driven techniques with MS-Excel, Wiley Publications

Reference Books

1. Stephan Sorger(2013).Marketing Analytics:Strategic models and metrics,Create Space Independent Publishing Platform
2. Chuck Hemann, Ken Burbary(2013). Digital Marketing Analytics, Pearson Publication
3. Rajkumar Venkatesan, Paul W.Farris, Ronald T.Wilcox(2021). Marketing Analytics: Essential Tools for Data-Driven Decisions, University of Virginia Press.
4. Robert W. Palmatier, J. Andrew Petersen , Frank Germann (2022) Marketing Analytics: Based on First Principles, Bloomsbury Academic

Pedagogy: Lectures, Demonstrations, Group Discussions, Case studies, Online References

Course Designers:

1. Dr. N. Radha
2. Mrs.G.Anitha

MDA23E3	Elective 1: TEXT ANALYTICS	Category	L	T	P	Credits
		Theory	58	2	-	4

Preamble

This course aims to introduce the principles of linguistics, natural language processing and text analytics. It focuses on models, techniques and tools for dealing with text data and to gain knowledge in recent advances in text analytics. It also provides text analytic framework and machine learning techniques for modelling the information in textual sources.

Prerequisite

- Statistics
- Data Structure
- Fundamentals of Information Retrieval

Course Learning Outcomes

On the successful completion of the course, students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Understand the basics, language syntax and structure of text analytics and NLP	K2
CLO2	Apply the concept of text preprocessing, tokenization and normalization	K3
CLO3	Analyze and parsing techniques, text classification, summarization and semantic relations	K4
CLO4	Evaluate classification models and clustering techniques	K5
CLO5	Build the model using text analytics techniques	K6

Mapping with Program Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	S	M	M
CLO2	S	S	S	M
CLO3	S	S	S	M
CLO4	S	S	S	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I (11 Hrs)

Introduction to Text Analytics: Introduction – Origin - Analytics in business and Industry -skills - benefits - **Text Analytics Process Road Map** - The Fundamentals of Content Analysis - **Planning for Text Analytics**

UNIT II (11 Hrs)

Natural Language Basics- Language Syntax and Structure – Language Semantics - Language as Data: A Computational Model – Language, Contextual & Structural Features – Text Corpora – Working with Text - String Operations and Methods - **Text Analytics Frameworks**

UNIT III (12 Hrs)

Processing and Understanding Text - Text Tokenization – Text Normalization – Understanding Text Syntax and Structure – **Parts of Speech Tagging** – Shallow parsing – **Dependency-based parsing – Constituency-based parsing**

UNIT IV (12 Hrs)

Text Classification - Feature Extraction – Classification Algorithms: **Multinomial Naïve Bayes** - Evaluating Classification Models – **Building a Multi-Class Classification System**-Application and uses – Text Summarization: Text Normalization – Feature Extraction – **Keyphrase Extraction** - Text Similarity: Analyzing Term Similarity - Clustering Greatest Movies of All Time: **K-Means Clustering**

UNIT V (12 Hrs)

Semantic and Sentiment Analysis: Semantic Analysis- Exploring WordNet – **Analyzing lexical semantic relations – Word sense disambiguation** – Named Entity Recognition – **Analyzing Semantic Representations** – Case Study: Sentiment Analysis of Movie Reviews

Text Books

1. Murugan Anandarajan, Chelsey Hill, Thomas Nolan (2016). Practical Text Analytics: Maximizing the Value of Text Data, Springer
2. Dipanjan Sarkar (2016). Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data (2/e), Apress Media LLC
3. David Robinson(2017), Text Mining with R: A Tidy Approach 1st Edition by Julia Silge, O'Reilly Media Inc.

Reference Books

1. Steven Bird, Ewan Klein, Edward Loper (2009). Natural Language Processing with Python. O'Reilly Media, Inc.
2. Benjamin Bengfort & Tony Ojeda & Rebecca Bilbro(2018), Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning, O'Reilly Media, Inc.

Pedagogy: Lectures, Demonstrations, Guest Lecture, Video Lectures

Course Designers

1. Ms.M.Nandhini
2. D. K. Gandhimathi

MDA23P3	DATA ANALYTICS LAB II	Category	L	T	P	Credit
		Practical	-	-	75	3

Preamble

This course provides implementation of statistics concepts in R and SPSS. It also provides sound introduction to implement the Hadoop framework. This course also provides various exercises to implement in the distributed environment through map reduce programming.

Prerequisite

- Statistics
- Big data framework
- Java

Course Learning Outcomes

On successful completion of the course, the students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Demonstrate concepts of probability in R	K2
CLO2	Apply different probability distributions in SPSS	K3
CLO3	Apply specific Mapper and reducer functions for different situations	K3
CLO4	Design and build a hadoop cluster. Create NoSQL databases using MongoDB.	K6
CLO5	Develop DML,DDL commands in Hbase.	K6

Mapping with Program Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	S	M	M
CLO2	S	S	M	M
CLO3	S	S	M	M
CLO4	S	S	S	M
CLO5	S	S	S	M

S- Strong; M- Medium; L- Low

Syllabus

- Setting up a hadoop environment
- Exercises to implement file management tasks using Hadoop
- Exercises to implement Map reduce program
- Exercises to implement Stock count Map reduce program.
- Exercises to implement CRUD, Aggregating and indexing operations in MongoDB.
- Exercises to apply the concept of MapReduce in MongoDB .
- Exercises to implement DDL, DML commands using HBase.
- Exercises to implement concepts of probability and distributions in R
- Exercises to implement concepts of probability and distributions in SPSS

Pedagogy: Demonstrations

Course Designers:

1. Dr. S.Meera
2. Dr. T.A. Albinaa

MDA23P4	MACHINE LEARNING LAB	Category	L	T	P	Credits
		Practical	-	-	75	3

Preamble

This course introduces Sci-kit learn, the popular machine learning libraries. It also provides exercises to implement Sci-kit learn syntax and tools to apply machine learning algorithms.

Prerequisite

- Knowledge in Open source - Python
- Machine Learning

Course Learning Outcomes

On successful completion of the course, students will be able to

CLO Number	CLO Statement	Knowledge Level
CLO1	Apply Machine Learning concepts to real time applications.	K3
CLO2	Build models using classification algorithm for real world problems	K6
CLO3	Build models using clustering algorithm for real world problems	K6
CLO4	Create classification and clustering models	K6
CLO5	Test and evaluate the models	K3

Mapping with Program Outcomes

CLOs	PO1	PO2	PO3	PO4
CLO1	S	S	S	M
CLO2	S	S	S	M
CLO3	S	S	S	M
CLO4	S	S	S	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

- Exercises to load dataset into sci-kit learn
- Exercise for Building models in sci-kit learn
- Exercise to extract features from datasets
- Exercise to implement Regression
- Exercise to implement Classification
- Exercise to implement Clustering
- Exercises for Model selection and evaluation
- Exercises to Build a data pipeline

Pedagogy: Demonstrations

Course Designer:

1. Mrs.G..Anitha
2. Dr.N.Radha

