PSGR
Krishnammal College for Women

College of Excellence, nirf 2021-6th 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
2022 – 2023 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.

PSGR Krishnammal College for Women

College of Excellence, nirf 2021-6th 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 – 2022-2023 BATCH ONWARDS
SEMESTER I

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	MDA2201	Paper 1: Descriptive Statistics	4	56	4	3	50	50	100	4
I	III	MDA2202	Paper 2: Foundations of Data Science	4	56	4	3	50	50	100	4
I	III	MDA2203	Paper 3: Linear Algebra	4	56	4	3	50	50	100	4
I	III	MDA2204	Paper 4: Data Structures	4	56	4	3	50	50	100	4
I	III	MDA2205	Paper 5: RDBMS and SQL	4	56	4	3	50	50	100	4
I	III	MDA22P1	Lab1: Oracle and SQL Lab	5	75	-	3	50	50	100	3
I	III	MDA22P2	Lab2: Data Analytics Lab I (R, SPSS, SciLab)	5	75	-	3	50	50	100	3
I	III		Online Course	-	-	-	-	-	-	-	-

SEMESTER I

MDA2201	DESCRIPTIVE STATISTICS	Category	L	T	P	Credit
		Theory	56	4	-	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.

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.

UNIT II **(11 Hrs)**

Presentation of data: Construction of Tables with one or more factors of classification, 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. Case Studies using Statistical Methods.

UNIT III **(11 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.

UNIT IV **(11 Hrs)**

Bivariate data – scatter diagram, correlation coefficient and its properties, Correlation ratio, Intraclass correlation, Rank correlation – Spearman's and Kendall's measures

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 – standard error of estimate-
Use Cases

Text Book

1. S.C.Gupta ,V.K.Kapoor (2019), Fundamentals of Mathematical Statistics, Sultan Chand & Sons (Unit I : Chapter 1, Unit IV: Chapter 10: 10.1 – 10.3, 10.7.1 -10.7.3, Chapter 12: 12.1 – 12.2)
2. S.P.Gupta (2021), Statistical Methods, Sultan Chand and Sons (Unit II: Chapter 5 (108-126), Chapter 6, Unit V: Chapter 11)
3. R.Wilcox (2009), Basic Statistics, Oxford University Press, (Unit III: 2.1-2.5)
4. Price Michael (2012) ,Excel 2010 In Easy Steps, Tata MCGraw - Hill (Unit II : Chapter 8 (123 -133,135))

Reference Books

1. U.Dinesh Kumar, Business Analytics,Wiley, 2017.
2. Murray R Spiegel and Larry J Stephens: Statistics, Schaum's Outline, Fourth edition, 2008
3. R.S.N. Pillai, Statistics, S. Chand Publishing Company Pvt Ltd , 1992
4. https://www.indiabix.com/data-interpretation/questions-and-answers/
5. https://www.mathsisfun.com/data/pictographs.html

Pedagogy

Lectures, Simulation exercises, Demonstration

Course Designers

1. Dr.T.A.Albinaa
2. Dr.C.R.Parvathy

MDA2202	FOUNDATIONS OF DATA SCIENCE	Category	L	T	P	Credit
		Theory	56	4	-	4

Preamble

This course provides the fundamental concepts in data science. It includes Data Classification, Sources of Data, Data Science user- roles and skills, Process of big data technology, Security and Intelligence, Basics of R and statistical measures.

Prerequisite

Mathematics / Statistics and Java

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 myths in big data, functionalities of R	K5
CLO5	Create solutions for the problems related to data science using R.	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 **(11 Hrs)**

Data Evolution: Data Development Time Line - ICT Advancement - a Perspective –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- Data Evolution - Data Sources - Statistical Inferences from Data.

UNIT II **(12 Hrs)**

Data Science: 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- Relation: Data Science, Analytics and Big Data Analytics. Data Science Components – Big data technology – Data Science user- roles and skills- Data Science use cases

UNIT III **(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, Big data Technology: Big Data Technology Process –Big Data Exploration -Data Augmentation – Operational Analysis –360 View of Customers –Security and Intelligence

UNIT IV **(11 Hrs)**

R Basics: Introduction- Packages and Library – Data types – Basic operators – R objects- Vectors – Lists- Arrays – Matrix- Factors – Data frame- R file formats- Importing and exporting files – Data Visualization in R: Lattice package- Box plot- bar chart – scatter plot- GGplot2

UNIT V **(11 Hrs)**

Statistical Measures in R: Measures of central tendency – Range- inter quartile range – Mean – Median – variance- Standard deviation – Sampling distribution – probability distributions- hypothesis tests – Time Series Analysis: Multivariate Time Series.

Use cases: Insurance policy offers, Discount targeting in online shopping

Text Book

1. V. Bhuvaneswari, T. Devi, (2016). Big Data Analytics: A Practitioner's Approach, Bharathiar University
2. V. Bhuvaneswari (2016). "Data Analytics with R – Step by Step", Bharathiar University.

Reference Books

1. Nina Zumal, John Mount (2014). Practical Data science in R, Managing Publication Company
2. Bernard Kolman, Robert C. Busby and Sharon Ross (2004). Discrete Mathematical Structures, New Delhi: Prentice Hall

Pedagogy: Lectures, Demonstration and Case Studies

Course Designers

1. Dr.G. Dona Rashmi
2. Dr.S. Poongodi

MDA2203	LINEAR ALGEBRA	Category	L	T	P	Credit
		Theory	56	4	-	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

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 **(11 Hrs)**

The Geometry of Linear Equations- An Example of Gaussian Elimination- Matrix Notation and Matrix Multiplication - Triangular Factors and Row Exchanges- Inverses and Transposes

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

UNIT III **(11 Hrs)**

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

UNIT IV **(11 Hrs)**

Eigenvalues and Eigenvectors: Introduction- Diagonalization of a Matrix .- Difference Equations and Powers A^k- Differential Equations and e^{At} - Complex Matrices- Similarity Transformations

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.

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

MDA2204	DATA STRUCTURES	Category	L	T	P	Credit
		Theory	56	4	-	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	M	M
CLO5	S	S	S	M

S- Strong; M-Medium; L-Low

Syllabus

UNIT I **(11 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 **(11 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, Sanguthevar Rajasekaran (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. Sartaj Sahni (2010). Data Structures, Algorithms, and Applications in C++; McGraw-Hill International Edition
3. AlfredV.Aho, John E.Hopcroft, Jeffrey D.Ullman (2000). Data structures and algorithms; Pearson Education, Asia.
4. Adam Drozdek (2013). Data Structures and Algorithm in C++ , 4th Edition.

Pedagogy: Lectures, Group Discussion, case study

Course Designers

1. Dr.G. Dona Rashmi
2. Dr. N. Radha

MDA2205	RDBMS AND SQL	Category	L	T	P	Credit
		Theory	56	4	-	4

Preamble

This course presents the fundamental concepts of Database Management Systems, database design and Relational model. It provides concepts of how to apply these in practice and learn how to use the structured query language to work and analyse databases.

Prerequisite

- Database
- Programming concepts

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 relational databases, database design Using ER diagram	K2
CLO2	Demonstrate various SQL queries by applying RDBMS concepts	K3
CLO3	Analyze various real time applications for applying RDBMS concepts.	K4
CLO4	Evaluate constraints on data and identify situations to apply the constraints.	K5
CLO5	Design different databases for various situations	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

Syllabus

UNIT I: **(11 Hrs)**

Overview of database systems: Managing Data - A Historical Perspective - File System versus DBMS – Advantages of a DBMS - Describing and Storing data in a DBMS - Queries in a DBMS - Transaction Management - Structure of a DBMS.

UNIT II: **(11 Hrs)**

Database Design & ER diagrams - Entities, Attributes and Entity Sets - Relationships and Relationship Set - Additional features of the ER model - Conceptual Database design with ER Model.

UNIT III: **(12 Hrs)**

Relational Model: Introduction - Integrity Constraints Over Relations - Enforcing Integrity Constraints on Relational Data - Logical Database Design: ER to Relational - Introduction to Views - Destroying / Altering Tables and Views - Relational Algebra and Calculus.

UNIT IV: **(11 Hrs)**

SQL Queries, Constraints, Triggers: The form of a Basic SQL Query - UNION, INTERSECT and EXCEPT - Nested Queries - Aggregate Operators - Null Values - Complex integrity constraints in SQL - Triggers and Active Data bases – Query Evaluation.

UNIT V: **(11 Hrs)**

PL/SQL Programming: Functions and Procedures, Triggers, Queries, Forms, Reports, Cursors, Exceptions. Introduction to NoSQL – Types.

Text Book

Raghu Ramakrishnan and Johannes Gehrke (2014). Database Management System, Third edition, McGraw-Hill.

Reference Books

1. Abraham Silberschatz, Henry F. Korth and Sudarshan S(2021). Database System Concepts, 7/e, McGraw-Hill Education (India)
2. Date CJ (2006). An Introduction to Database Systems,8/e, Pearson Education.
3. Michael Mclaughlin, (2017).Oracle Database 11g PL/SQL Programming, McGraw Hill.
4. Shashank Tiwari (2011). Professional NoSQL, John Wiley & Sons

Pedagogy: Lectures, Group Discussion, Demonstration

Course Designers:

1. Dr. N. Radha
2. Dr.S. Poongodi

MDA22P1	ORACLE AND SQL LAB	Category	L	T	P	Credit
		Theory	-	-	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 integrity constraints on databases, functions, procedures, cursors, triggers, exception handling, forms and reports.

Prerequisite

- Database concepts
- Programming concepts

Course Learning Outcomes

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

CLO Number	CLO Statement	Knowledge Level
CLO1	Implement the databases concepts and SQL queries as per implementation.	**K2**
CLO2	Apply specific SQL commands on relational tables for different situations	**K3**
CLO3	Analyse use cases and create constraints suitable for the given situation.	**K4**
CLO4	Create and analyse a database using SQL DML/DDL commands	**K6**
CLO5	Design and build a GUI application	**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

Syllabus

- Exercises to implement the concepts of null constraint, unique constraint, integrity constraints, check 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
- Exercise to create forms and reports.
- Exercises to create cursors and exceptions
- Exercises to create Simple applications

Course Designers:

1. Dr. N. Radha
2. Dr. S. Poongodi

MDA22P2	DATA ANALYTICS LAB I

Category	L	T	P	Credit
Practical	-	-	75	3

Preamble

The course deals with the implementation of linear algebra concepts in Scilab. This course also provides various exercises to implement the statistical functions using R and SPSS.

Prerequisite

- Statistical concepts
- Basic concepts of 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 basic concepts of Linear algebra and statistics	**K2**
CLO2	Implement the algebraic and statistical problems using Scilab and R	**K3**
CLO3	Apply the concepts of Linear algebra and statistics in real time problems	**K4**
CLO4	Analyse real time data using various statistical measures in scilab	**K5**
CLO5	Construct models using various statistical methods in R and SPSS	**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

Syllabus

- Exercises to implement the basic matrix operations in Scilab.
- Exercises to find the Eigenvalues and eigenvectors in Scilab.
- Exercises to solve equations by Gauss elimination, Gauss Jordan Method and Gauss Siedel in Scilab.

- Exercises to implement the associative, commutative and distributive property in a matrix in Scilab.
- Exercises to find the reduced row echelon form of a matrix in Scilab.
- Exercises to plot the functions and to find its first and second derivatives in Scilab.
- Exercises to present the data as a frequency table in SPSS.
- Exercises to find the outliers in a dataset in SPSS.
- Exercises to find the most risky project out of two mutually exclusive projects in SPSS
- Exercises to draw a scatter diagram, residual plots, outliers leverage and influential data points in R
- Exercises to calculate correlation using R
- Exercises to implement Time series Analysis using R.
- Exercises to implement linear regression using R.
- Exercise to perform Exploratory Data Analysis in Excel for applications like,
    - Service Tax Collection
    - Electricity Consumption
    - Sales Gross Profit and Net Profit
    - Road deaths in a City for a Specific Period
    - Percentage of People who report paying bribe in the previous 12 months.

Course Designers

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