Academic Council: -/--/--
Item No. :

College of Arts, Science & Commerce
SIES
RISE WITH EDUCATION
Sion (West), Mumbai – 400022.
(Autonomous)

Faculty: Science
Program: B.Sc.
Subject: Data Science
Academic year: 2024-2025

Syllabus for Semester- III and Semester – IV

Choice Based Credit System Syllabi (as per NEP) approved by Board of Studies in Data Science to be brought into effect from June 2024.

PREAMBLE

Data has become the most important factor in this era of digital transformation. The technological advancements are seen in all walks of life and therefore we are flooded with enormous data. Every business relies on data to deliver better products as well as services. All data are stored in cloud, and so accessed and processed easily. Data analytics has helped in better decision making with sufficient data insights.

Predictive Analysis has played a crucial role in making businesses smarter with improvised strategies. Machine Learning and Artificial Intelligence are used together to optimize business operations and data management. Augmented analytics uses machine learning and natural language processing to automate the process of data analysis. Global data is predicted to grow due to data generated by the Internet of Things(IoT) and cloud computing advancements. These developments have given rise to a new area of study, called Data Science.

Data Science as an area has evolved out of the applications of various tools and techniques in the field of Computer Science, Mathematics and Statistics. There is an increasing demand to capture, analyse the enormous data present in a number of application domains. The data in these applications then needs to be converted into actionable strategies for effective decision making. So, the study of data science has become essential to cater to the growing need of data scientists and data analysts.

This course focuses on educating the students about the essentials of computer science, applied mathematics and applied statistics with respect to the data science applications.

PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOME

NO.	Details
PO1.	**Solving Complex Problems**:- Apply the knowledge gained in breaking down complex problems into simple components; and to design processes required for problem solving.
PO2.	**Critical Thinking: -** Ability to apply the acquired knowledge to identify assumptions and evaluate their accuracy and validity.
PO3.	**Reasoning ability and Rational thinking: -** Ability to analyse, interpret data and draw logical conclusions; to evaluate ideas rationally.
PO4.	**Research Aptitude: -** Ability to ask relevant questions to identify and define the problem, applying research tools for analysis and interpretation of data. Understand and comply with research ethics.
PO5.	**Effective Communication skill: -** Demonstrate the ability to listen and to clearly express ideas verbally. Equip to write reports, make presentations effectively.
PO6.	**Information and Digital Literacy: -** Equip to use appropriate tools and techniques inclusive of internet and electronic media for acquiring, assessing and analysing data from diverse resources.
PO7.	**Social Interactive Skills and team work: -** Exhibit networking and social interactive skills; function effectively as an individual and as a member in diverse groups; demonstrate leadership quality useful for employability
PO8.	**Self-directed and Lifelong Learning:** Ability to explore and gain knowledge in independent and self-reliant ways. Demonstrate ability to adapt and upgrade with the global , social and technological changes.
PSO1.	**Sound Knowledge:** Demonstrate the knowledge of core data science concepts and apply them to develop a user- friendly, scalable, and robust applications
PSO2.	**Critical and Rational Thinking:** Exhibit higher order skills to adapt to the everchanging technological environment
PSO3.	**Logic Building and Programming Skills:** The ability to apply logic to problem solving and acquire proficiency in various programming languages.
PSO4.	**Data Analysis :** Apply quantitative modelling and data analysis techniques to solve real world business problems, Learn tools and techniques for transformation of data and statistical data analysis
PSO5.	**Pursue Higher Education:** Make students competent to take up advanced degree courses like MSc(Data Science),MCA, MSc(CS), MSc(IT) and MBA etc.

S.Y.B.Sc.(DS) under NEP - Semester III

Course Code	Course Type	Course Title	Credits
SIUDSMJ211	Core Subject (Major)	Data Science with R	**3**
SIUDSMJ212	Core Subject (Major)	Probability and Distributions	**3**
SIUDSMN211	Core Subject (Minor)	Database Technology – I	**3**
SIUDSMJP211	Core Subject (Major) Practical	Data Science with R – Practical	**1**
SIUDSMJP212	Core Subject (Major) Practical	Probability and Distributions – Practical	**1**
SIUDSMNP211	Core Subject (Minor) Practical	Database Technology – I – Practical	**1**

SIUDSMJ211 : Data Science with R

B.Sc. (Data Science)	Semester – III
Course Name: Data Science with R	**Course Code: SIUDSMJ211**
Credits	**3**

Course Objective:

- Master the use of the R interactive environment and expanding by installing R packages
- Read Structured Data into R from various sources
- Understand the different data types and data structures in R
- Manipulate strings, dates in R
- Understand basic regular expressions in R
- Understand base R graphics
- Focus on GGplot2 graphics for R and be familiar with trellis (lattice) graphics.

Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** To use R Studio and explore the features for R programming
- **CO2:** To use R functions and graphics with in R programming for solving problems.
- **CO3:** To work with advanced graphics of R, import and use the data and represent the data into tables.
- **CO4:** To manipulating Data Frames and make use of Dates in R application.

Unit	Contents	No. of Lectures
I	**Getting started with R:** R Software: Obtaining R and RStudio, First R Encounter, Getting started: R as a big calculator, Assignment, Basic operators, Help with functions and features **R Interfaces**: Using R and RStudio: R Software, Obtaining R and RStudio, The default R interface, RStudio Interface, Example Datasets in R, R Packages, Installing new R libraries, Customizing R Start-up **Objects in R:** Using ls and rm to managing R Objects, Types of R objects, Attributes of R Objects, Creating and accessing objects, Modifying elements. **Reading and writing data to and from R:** Importing and reading text files data into RStudio, Importing data using R command read.table(), Importing text files Using scan(), Parsing each line – Readlines, Writing Data table from R, Importing Data from other Software, Reading data from Excel into R, Import/Export from other statistical software, From a Database Connection, Sampling and Creating simulated data. **Introduction to programming and writing Functions in R:** Functions, Conditional statements (if, ifelse, switch), Repetitive execution: For and While loops	15

II	**R Apply Package:** The Apply Functions, Functions for parsing text, Programming in R: More advanced, Viewing Code of functions from R packages, Parsing Real Data - World Population Data from Wikipedia, Writing functions: more technical discussion -Scoping, Options for Running memory or CPU intensive jobs in R, Efficient R coding **Importing Data – readr:** Functions for Reading Data, File Headers, Column Types, String-based Column Type Specification, Function based Column Type Specification Parsing Time and Dates, Space-separated Columns, Functions for Writing Data **Introduction to graphics in R:** The R function plot(), Customize plot with low-level plotting commands, Default parameters – par, Interacting with graphics, Saving plots, Useful Graphics Resources **Advanced Graphics:** Advanced plotting using Trellis; ggplots2, Lattice, Examples that Present Panels of Scatterplots using xyplot(), Simple use of xyplot	15
III	**Reformatting Tables – tidyr:** Tidy Data, Gather and Spread, Complex Column Encodings, Expanding, Crossing, and Completing, Missing Values, Nesting **Data Pipelines – magrittr:** The Problem with Pipelines, Pipeline Notation, Pipelines and Function Arguments, Function Composition, Other Pipe Operations **Working with Strings – stringr:** Counting String Patterns, Splitting Strings, Capitalizing Strings, Wrapping, Padding, and Trimming, Detecting Substrings, Extracting Substrings, Transforming Strings **Working with Factors – forcats:** Creating Factors, Concatenation, Projection, Adding Levels, Reorder Levels **Manipulating Data Frames – dplyr:** Selecting Columns, Filter, Sorting, Modifying Data Frames, Grouping and Summarizing, Joining Tables. **Working with Dates – lubridate**: Time Points, Time Zones, Time Intervals	15

Books and Reference

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	R Programming for Data Science	Roger D Peng		1st	2015
2	Data Science from Scratch	Joel Grus	O'Reilly Media, Inc.	2nd	2019
3	An Introduction to Statistical Learning	Gareth James, Daniela Witten, Trevor Hastie, Robert	Springer Science & Business Media, 2013	Illustrated	2013

		Tibshirani			
4	Practical Statistics for Data Scientists	Peter Bruce, Andrew Bruce	O'Reilly Media, Inc.	3rd	2018

SIUDSMJP211 : Data Science with R Practical

B.Sc. (Data Science)	Semester – III
Course Name: Data Science with R Practical	**Course Code: SIUDSMJP211**
Credits	**1**

List of Practical:

1	**Introduction to R Programming Elements**
a.	Write an R Program to implement expressions, assignment and decision making
b.	Write an R Program to design and implement loops.
c.	Write a R program to demonstrate the use of essential data structures in R [Hint: Vectors, Matrix, Arrays]
2	**Using List, Data Frames and Functions in R**
a.	Write an R program to manage data and exhibit operations on it using List data structure
b.	Write an R program to manage data and exhibit operations on it using Data Frames
c.	Write an R program to demonstrate the use of: User-defined functions Built-in numeric function, character functions etc.
3	**Implement decision making in R**
a.	Write an R program to implement if, if-else, if-else-if ladder, nested if-else, and switch
4	**Implementing Strings in R**
a.	Write an R program to store and access string in R objects(vectors, matrix, arrays, data frames, and lists)
b.	Write an R program to demonstrate use of various string manipulation functions. [Hint: paste(), print(), noquote(),format(), cat(), toString(), sprint()]
5	**Performing Statistics with R-I**
a.	Write an R program to apply built-in statistical functions. [Hint: mean, median, standard deviation and others]
b.	Write an R program to demonstrate Linear and Multiple Regression analysis.
6	**Performing Statistics with R-II**
a.	Write an R program to implement Normal Distribution. [Hint: dnorm(), pnorm(), qnorm(), rnorm()] Binomial Distribution: [Hint: dbinom(), pbinom(),qbinom(),rbinom()]
7	**Data Visualization and Analysis**
a.	Write an R program to demonstrate various ways of performing Graphical analysis.
8	**Data Interfaces in R**
a.	Write an R program to demonstrate data interface with CSV files [Hint: creating data for CSV, analyzing, writing CSV files]
b.	Write an R program to work with spreadsheet (Excel) programs. [Hint: installing, loading, verifying, creating data for xlsx file]
c.	Write an R program to manage data using XML files. [Develop data interface for

	maintaining Employee Information]
9	**Handling Errors in R**
a.	Write an R program to demonstrate various error messages in R Programming
b.	Write an R program to implement Error Handling in R [Hint: warning(),stop(),try(), tryCatch(), CallingHandlers()]
10	**Handling Dates in R**
a.	Write R program to demonstrate dates in R.

SIUDSMJ212 : Probability and Distributions

B.Sc. (Data Science)	Semester – III
Course Name: Probability and Distributions	**Course Code: SIUDSMJ212**
Credits	**3**

Course Objective:

- To explore about random variables and implement various distribution functions
- To familiarize with concepts of probability and learn implementation of different types of probabilities.
- Learn and implement the concept of expectation, related theorems and generating functions.
- To know the concept and implementation of discrete distributions including Bernoulli, Binomial and power series distributions.

Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Organize, manage and present data.
- **CO2:** Analyse statistical data graphically using frequency distributions and cumulative frequency distributions.
- **CO3:** Use the basic probability rules, including additive and multiplicative laws, using the terms independent and mutually exclusive events.
- **CO4:** Derive the probability density function of the transformation of random variables.
- **CO5:** Calculate probabilities and derive the marginal and conditional distributions of bivariate random variables.

Unit	Contents	No. of Lectures
I	**Theory of Probability:** Introduction, history, different terms, mathematical tools, Axiomatic approach to probability, Mathematical notation, multiplication and conditional probability, Bayes theorem, Geometric probability. **Random Variables and Distribution Functions:** Random Variable, distribution function, discrete random variable, continuous random variable, joint probability law, transformation of one-dimensional random variable, transformation of two-dimensional random variable	15
II	**Mathematical Expectation and Generating Functions:** Mathematical expectation, Expectation of a Function of a Random Variable, Addition Theorem of Expectation, Multiplication Theorem of Expectation, Expectation of a Linear Combination of Random Variables, Covariance, Variance of a Linear Combination of Random Variables, Moments of Bivariate Probability Distributions, Conditional Expectation, and Conditional Variance, Moment Generating Function, Cumulants, Characteristic Function, Chebychev's Inequality, Convergence in- Probability, Weak Law of	15

	Large Numbers, Borel Canteli Lemma, Probability Generating Function.	
III	**Theoretical Discrete Distributions:** Introduction, Bernoulli distribution, Binomial Distribution, Poisson Distribution, Negative Binomial Distribution, Geometric Distribution, Hypergeometric Distribution, Multinomial Distribution, Discrete Uniform Distribution, Power Series Distribution **Theoretical Continuous Distributions:** Rectangular or Uniform Distribution, Normal Distribution, Gamma Distribution, Beta Distribution of First Kind, Beta Distribution of Second Kind, The Exponential Distribution, Laplace Double Exponential Distribution, Weibul Distribution, Cauchy Distribution, Central Limit Theorem, Compound Distributions, Pearson\s Distributions, Variate Transformations, Order Statistics, Truncated Distributions.	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Fundamentals of Mathematical Statistics	S.C, Gupta, and V. K. Kapoor	S. Chand and Sons	10^{th}	2002
2	Applied Statistics and Probability for Engineers	Douglas C. Montgomery and George C. Runger	Wiley	6^{th}	2014
3	Probability, Statistics, and Stochastic Processes	Peter Olofsson And Mikael Andersson	Wiley	2^{nd}	2012

SIUDSMJP212 : Probability and Distributions Practical

B.Sc. (Data Science)	Semester – III
Course Name: Probability and Distributions Practical	**Course Code: SIUDSMJP212**
Credits	**1**

List of Practical: (Can be done in MS-Excel)

1	**Introduction to Probability**
a.	Formulate and apply Bayes' Theorem Calculations for problems like the "Two Supplier Example". [Hint: Use Prior Probabilities and Conditional Probabilities to compute Joint and posterior probabilities.]
b.	Design a spreadsheet to demonstrate the association Between Two Variables by Computing the Covariance and Correlation Coefficient. [Hint: Use COVAR and CORREL)]
2	**Discover Probability using formulas**
a.	Design and spreadsheet experiment to compute the probability using the geometric distribution formula.
b.	Create a spreadsheet application to compute the Conditional Probability. Also determine the probability that a randomly chosen event.
3	**Random Variables and Distribution Functions**
a.	Create a spreadsheet application to Compute the Expected Value, Variance, and Standard Deviation
b.	Create a spreadsheet application to Compute Binomial Probabilities. [Hint: Use BINOM DIST]
4	**Probability Distribution and Law**
a.	Create a spreadsheet application to Poisson Probability Distribution. [Hint: Use POISSON]
b.	Create a spreadsheet application to implement joint probability law.
5	**Mathematical Expectation and Chebychev's Theorem**
a.	Create a spreadsheet application to compute the expectation of a Function of a Random Variable
b.	Create a spreadsheet application to apply Chebychev's Theorem
6	**Conditional Expectation and Generating Functions**
a.	Create a spreadsheet application to compute Conditional Expectation and Conditional Variance.
b.	Create a spreadsheet application to demonstrate the use of Generating Functions
7	**Theoretical Discrete Distributions1**
a.	Create a spreadsheet application to demonstrate Bernoulli Distribution.
b.	Create spreadsheet application to use excel function for computing hypergeometric probabilities.
8	**Theoretical Discrete Distributions2**

a.	Create spreadsheet application to Calculate Binomial Distribution in Excel. [Hint: Use BINOM.DIST]
b.	Create suitable spreadsheet application to work with PowerSeries Distribution.
9	**Theoretical Continuous Distributions1**
a.	Create spreadsheet application for computing probabilities and zvalues for the standard normal distribution. [Hint: Use NORMSDIST and NORMSINV]
b.	Create spreadsheet application for computing probabilities for the exponential probability distribution. [Hint: Use EXPONDIST]
10	**Theoretical Continuous Distributions2**
a.	Create spreadsheet application for demonstrating Weibull Distribution to obtain a model for data sets. [Hint: Use WEIBULL.DIST]
b.	Create spreadsheet application for demonstrating Pearson\s Distributions.

SIUDSMN211: Database Technology-I

B.Sc. (Data Science)	Semester – III
Course Name: Database Technology-I	**Course Code: SIUDSMN211**
Credits	**3**

Course Objective:

- This course concentrates the concept of the DBMS with respect To principles, design and implementation of DBMS. It aims to specify the functional and Data requirements for a typical database application and to understand creation, manipulation And querying of data in databases.
- To understand Organizing, structuring and storing data
- Understand Database as Relational model
- To understand SQL to retrieve data and concept of redundancy
- To specify the functional and data requirements for a typical database application
- To understand creation, manipulation and querying of data in databases

Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Students should be able to evaluate business information problem and find the requirements of a problem in terms of data.
- **CO2:** Students should be able to draw database design in logical structure and can identify the entities which exist in a system
- **CO3**: Students should be able to construct normalized database and functional dependencies between attributes and relational algebra queries
- **CO4:** Students should be able to design the database schema with the use of appropriate datatypes for storage of data in database.
- **CO5:** Students should be able to create, manipulate, query and backup the databases with features of SQL.

Unit	Contents	No. of Lectures
I	**Introduction to Databases and Transactions** What is database system, purpose of database system, view of data, relational databases, database architecture, transaction management **Data Models** The importance of data models, Basic building blocks, Business rules,The evolution of data models, Degrees of data abstraction. **Database Design and ER Diagram** Database design and ER Model: overview, ER Model, Constraints, ER Diagrams, ERD Issues, weak entity sets, Codd's rules, Relational Schemas **Relational database model:** Logical view of data, keys, integrity rules, Relational Database design: features of good relational database design, atomic domain and Normalization (1NF, 2NF, 3NF, BCNF). **Relational Algebra**	15

II	**Structured Query Language(SQL):**Introduction to SQL, Basic SELECTQueries, SELECT Statement Options, FROM Clause Options, ORDERBY Clause Options, WHERE Clause Options, Aggregate Processing, Subqueries, SQL Functions, Relational Set Operators, Crafting SELECT Queries	15
III	**Constraints, Views and SQL** Constraints, types of constrains, Integrity constraints, Views: Introduction to views, data independence, security, updates on views, comparison between tables and views SQL: data definition, aggregate function, Null Values, nested sub queries, Joined relations. Triggers **Transaction management and Concurrency Control** Transaction management: ACID properties, serializability and concurrency control, Lock based concurrency control (2PL, Deadlocks), Time stamping methods, optimistic methods, database recovery management.	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Database System and Concepts	A Silberschatz, H Korth, SSudarshan	Mc Graw Hill	5^{th}	
2	Database Systems	Rob Coronel	Cengae Learning	12^{th}	
3	Introduction to Database System	C.J.Date	Pearson	1^{st}	2003

SIUDSMNP211 : Database Technology-I Practical

B.Sc. (Data Science)	Semester – III
Course Name: Database Technology-I Practical	**Course Code: SIUDSMNP211**
Credits	**1**

List of Practical:

1	**Create ER-diagram for Hospital and library system.**
2	**Draw E-R diagram and convert entities it to table.**
3	**Write SQL query for given problem statement**
a.	Creating a Database
b.	Viewing all databases
c.	Viewing all Tables in a Database
4	**Perform the following:**
a.	Create table
b.	Drop table
c.	Truncate
5	**Manipulating Data**
a.	Using INSERT statement
b.	Using DELETE statement
c.	Using UPDATE statement
6	**Creating and Managing Tables**
a.	Creating and Managing Tables
b.	Including Constraints
7	**Perform following:**
a.	Simple Queries with Where Operators
b.	Where with Keywords and Logical Operators
c.	Queries with Aggregate functions (group by and having clause)
8	**Retrieving Data from Multiple Table:**
a.	Joining Tables (Inner-Joins, Outer-Joins)
b.	Aliases for Table Names
9	**Subqueries:**
a.	With IN clause
b.	With EXISTS clause
10	**Continue with the Employee Database and where each record consists of the following attribute**

	Attributes	Data Type	Description
1.	Pay_scale	Number	Pay scale of employee
2.	Basic_pay	Number	Employee's basic pay
3.	Net Pay	Number	Employee's deduction
4.	Gross Pay	Number	Employee's gross pay

Set emp_Id as primary key
Set the validation rule for basic_pay as (>8000) and the validation text as "Basic pay shouldbe greater than 8000"
Set the caption value of the field "Emp_Name" as "Employee Name"
Set the field "pay_scale" as a required(mandatory) fieldv)
Enter 10 records

S.Y.B.Sc.(DS) under NEP - Semester IV

Course Code	Course Type	Course Title	Credits
SIUDSMJ221	Core Subject (Major)	Big Data Analytics through Spark	**3**
SIUDSMJ222	Core Subject (Major)	Matrix Theory and Linear Algebra	**3**
SIUDSMN221	Core Subject (Minor)	Database Technology – II	**3**
SIUDSMJP221	Core Subject (Major) Practical	Big Data Analytics through Spark – Practical	**1**
SIUDSMJP222	Core Subject (Major) Practical	Matrix Theory and Linear Algebra – Practical	**1**
SIUDSMNP221	Core Subject (Minor) Practical	Database Technology – II – Practical	**1**

SIUDSMJ221 : Big Data Analytics through Spark

B.Sc. (Data Science)	Semester – IV
Course Name: Big Data Analytics through Spark	**Course Code: SIUDSMJ221**
Credits	**3**

Course Objective:

- Understand the Big Data Platform and its Use cases.
- Provide an overview of Apache Hadoop.
- Provide HDFS Concepts and Interfacing with HDFS.
- Understand Map Reduce Jobs Provide hands on Hadoop Eco System.
- Apply analytics on Structured, Unstructured Data.

Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Identify Big Data and its Business Implications.
- **CO2:** List the components of Hadoop and Hadoop Eco-System.
- **CO3:** Access and Process Data on Distributed File System.
- **CO4:** Manage Job Execution in Hadoop Environment.
- **CO5:** Develop Big Data Solutions using Hadoop Eco System.
- **CO6:** Analyze Infosphere BigInsights Big Data Recommendations.

Unit	Contents	No. of Lectures
I	**Big Data Technology Landscape:** Hadoop, HDFS, MapReduce, Hive **Apache Spark:** Evolution, Features, Spark Built on Hadoop, Components of Spark. **Spark architecture:** Resilient Distributed Dataset (RDD), Directed Acyclic Graph (DAG).	15
II	**Interactive Data Analysis with Spark Shell:** REPL Commands, Using the Spark Shell as a Scala Shell, Number Analysis, Log Analysis. **Writing a Spark Application:** Hello World in Spark, Compiling and Running the Application, Monitoring the Application, Debugging the Application. **Introducing Spark Streaming:** Spark Streaming Is a Spark Add-on, High-Level Architecture, Data Stream Sources, Receiver, Destinations, Application Programming Interface (API), Streaming Context, Basic Structure of a Spark Streaming Application, Discretized Stream (DStream), Creating a Dstream, Processing a Data Stream, Output Operations, Window Operation.	15

III	Introducing Spark SQL: Integration with Other Spark Libraries, Performance, Applications, ETL(Extract Transform Load), Application Programming Interface (API), Built-in Functions, Aggregate, Collection, Date/Time, Math, String, Window Introducing Machine Learning: Features, Labels, Models, Training Data, Test Data, Machine Learning Applications,	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Spark : The Definitive Guide	Bill Chambers & Mataei Zaharia	O'Reilly Media, Inc.	1st	2018
2	Big Data Analytics with Spark : A Practitioner's Guide to Using Spark for Large Scale Data Analysis	Mohammed Guller	Apress	1nd	2015
3	Practical Big Data Analytics Hands-on Techniques to Implement Enterprise Analytics and Machine Learning Using Hadoop, Spark, NoSQL and R	Nataraj Dasgupta	Packt	Illustrated	2018

SIUDSMJP221 : Big Data Analytics through Spark

B.Sc. (Data Science)	Semester – IV
Course Name: Big Data Analytics through Spark Practical	**Course Code: SIUDSMJP221**
Credits	**1**

List of Practical:

1.	Installation of Spark. Write a Hello World in Spark.
2.	Implement a program in Pig.
3.	Implement word count/ frequency program using MapReduce.
4.	Configure the Hive and implement the application in Hive.
5.	Implement Spark SQL.
6.	Implement machine learning with Spark or Hadoop.
7.	Implement Spark Streaming.
8.	Demonstrate Spark Shell commands.

SIUDSMJ222 : Matrix Theory and Linear Algebra

B.Sc. (Data Science)	Semester – IV
Course Name: Matrix Theory and Linear Algebra	**Course Code: SIUDSMJ222**
Credits	**3**

Course Objective:

- Understand the basic concepts of matrix algebra and its applications.
- Solving computational problems of linear algebra.

Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Understand basic matrix properties like rank, determinant, inverse, and a special type of matrices
- **CO2:** Introduce Gaussian / Gauss-Jordan elimination methods, LU factorization technique
- **CO3:** Use computational techniques for singular value decomposition (Computational and Algebraic Skills).
- **CO4:** Understand the concepts of vector space and subspaces.
- **CO5:** Understand the use of linear algebra and matrices in several important, modern applications of research and industrial problems involving statistics.

Unit	Contents	No. of Lectures
I	**Matrix theory:** Algebra of Matrices, Trace and Rank of a Matrix and their properties, Determinants, Inverse, Eigen values and Eigen vectors, symmetric, orthogonal and idempotent matrices and their properties **Matrix Factorization:** Gauss elimination, row canonical form, diagonal form, triangular form, Gauss-Jordan-LU decomposition, solving systems of linear equations. **Decomposition of Matrices:** Spectral decomposition, singular value decomposition, Quadratic forms, definiteness and related results with proofs.	15
II	**Vector Spaces:** Vector Spaces, Subspaces, Basis and dimension of a vector space, linear dependence and linear independence, spanning set. **Linear transformation:** Linear transformation, kernel, range, Matrix Representation of a linear transformation, rank- nullity theorem, change of basis and similar matrices. **Inner product spaces:** Inner-product spaces, orthogonal sets and bases, Orthogonal Projection, Gram-Schmidt orthogonalization process, Orthogonal Bases and Gram-Schmidt, The Fast Fourier Transform **Determinants:** Introduction, Properties of the Determinant, Formulas for the Determinant, Applications of Determinants	15

III	**Eigenvalues and Eigenvectors:** Introduction, Diagonalization of a Matrix, Difference Equations and Powers Ak, Differential Equations and eAt, Complex Matrices, Similarity Transformations **Computations with Matrices:** Introduction, Matrix Norm and Condition Number, Computation of Eigenvalues, Iterative Methods for Ax= b **Linear Programming and Game Theory:** Linear Inequalities, The Simplex Method, The Dual Problem, Network Models, Game Theory	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Introduction to linear algebra	Gilbert Strang	Wellesley-Cambridge	6th	2023
2	Linear Algebra and Its Applications	David C. Lay, Steven R. Lay, Judi J. McDonald	Pearson Education	5th	2023
3	Linear Algebra, Theory and Applications	Kenneth Kuttlet			2014
4	Basics of Matrix Algebra for Statistics with R	Fieller N	Chapman and Hall/CRC		2015

SIUDSMJP222: Matrix Theory and Linear Algebra Practical

B.Sc. (Data Science)	**Semester – IV**
Course Name: Matrix Theory and Linear Algebra Practical	**Course Code: SIUDSMJP222**
Credits	**1**

List of Practical: (use R/python/scilab/matlab)

1	**Matrices and Gaussian Elimination**.
a.	Multiplication and transpose of matrix.
b.	Inverses of matrix without using any inbuilt package.
c.	Inverses of matrix using any inbuilt package like numpy.
d.	Linear equation with n unknowns using Gauss Elimination
2	**Vector**
a.	Addition, subtraction, multiplication and division of vector
b.	Dot product & cross product of vector
c.	Visualizing vector Linear Transformations
3	**Orthogonal and Orthonormal Vectors**
a.	Computes the orthonormal vectors using the GS algorithm
b.	Projections and Least Squares
c.	Fast Fourier Transform
4	**Determinant of matrix**
a.	Finding determinant of matrix without using any inbuilt package.
b.	Finding determinant of matrix using any in built package.
5	**Calculate Eigenvalues**
a.	Compute the eigenvalues and right eigenvectors of a given square array
b.	Program to test diagonalizable matrix
6	i.Tests for Positive Definiteness ii.Singular Value Decomposition iii.The Finite Element Method (**Only Demonstration**)
7	Simplex Method. (**Only Demonstration**)
8	The Dual Problem. (**Only Demonstration**)
9	Implementing Network Models (**Only Demonstration**)
10	Implementing Game Theory (**Only Demonstration**)

SIUDSMN221 : Database Technology-II

B.Sc. (Data Science)	Semester – IV
Course Name: Database Technology-II	**Course Code: SIUDSMN221**
Credits	**3**

Course Objective:

- To understand the basics of PL/SQL.
- To understand control and conditional statement in PL/SQL.
- To understand working of sequences and cursor in PL/SQL.
- To understand concept of stored procedure and functions.
- To understand triggers and packages in PL/SQL.

Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Understand the basics of PL/SQL.
- **CO2:** Use of the control and conditional statement in PL/SQL.
- **CO3:** Apply sequences and cursor in PL/SQL.
- **CO4:** Know the concept of stored procedure and functions
- **CO5:** Create the triggers and packages in PL/SQL.

Unit	Contents	No. of Lectures
I	**Overview of Oracle Architecture:** Oracle Physical architecture and Oracle Instance Architecture. **The PL/SQL Block,** Lexical Units: Identifiers, Delimiters, Literals, Comments, Variables, PL/SQL Types, Expression, Operators and Control Structures, Records, **Cursors:** Definition of Cursor, Explicit and Implicit Cursors, Cursor for loops, Cursor variables, parametrized cursor	15
II	**Sub Program:** Procedures, Functions, Subprogram creation, parameter modes, Procedure versus Functions, **Database Triggers:** Use of Database Triggers, Types of Triggers, Creation and Deletion of Triggers, **Error Handling:** Declaring Exception, Raising Exception, Handling Exception, **Sequences & Pseudo Columns:** CURRVAL & NEXTVAL, ROWID, ROWNUM, **PL/SQL Transactions:** Transaction Control statements: commit, savepoint and rollback	15
III	**Locks:** Types of Locks and levels of locks. **Database Security:** Database Security Issues, Security Models, Different threats to databases, Challenges to maintaining database security, **Introduction to Other Databases:** Introduction to Parallel and distributed Databases, Introduction to Object Based Databases, XML Databases, NoSQL Database, Multimedia Databases, Big Data Databases	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Database System Concepts	Silberschatz A., Korth H., Sudarshan S.	McGraw Hill Education	6th	
2	SQL, PL/SQL the Programming Language of Oracle	Ivan Bayross	BPB Publications	4th	
3	Oracle PL/SQL Programming	Steven Feuerstein and Bill Pribyl	O'Reilly Publications	6th	2014
4	NoSQL Distilled A Brief Guide to the Emerging World of Polyglot Persistence	Pramod J. Sadalage, Martin Fowler, Addison-Wesley	Pearson Education, Inc.	1st	2012
5	Database Systems: Concepts, Design and Application	S.K.Singh	Pearson Education		

SIUDSMNP221: Database Technology-II Practical

B.Sc. (Data Science)	Semester – IV
Course Name: Database Technology-II Practical	**Course Code: SIUDSMNP221**
Credits	**1**

List of Practical:

1	**PL/SQL Basics**
a.	Use of variables.
b.	Write executable statement.
c.	Interacting with Oracle Server.
d.	Create anonymous PL/SQL block
2	**Control Structure in PL/SQL**
a.	Using while loop
b.	Do loop
c.	For loop
d.	Use of GOTO statement
3	**Create conditional statement using PL/SQL**
a.	Using if statement
b.	Using if else statement
c.	Using elsif ladder
d.	Using case expression.
4	**Creation of Sequence in PL/SQL**
5	**Create cursor in PL/SQL**
a.	Implicit cursor
b.	Explicit
c.	Parameterized cursor
d.	Cursor for loop
6	**Creation of Procedures in PL/SQL**
7	**Functions in PL/SQL**
a.	Compute and returns the maximum value
b.	Compute factorial of given number.
8	**Creation of Trigger**
a.	Create Row level trigger
b.	Create Statement level trigger
c.	Create instead of trigger
9	Handling exceptions
a.	Creation of user defined exception
b.	Creation of system defined exception
10	Design and Develop Application for Student Mark Sheet Processing