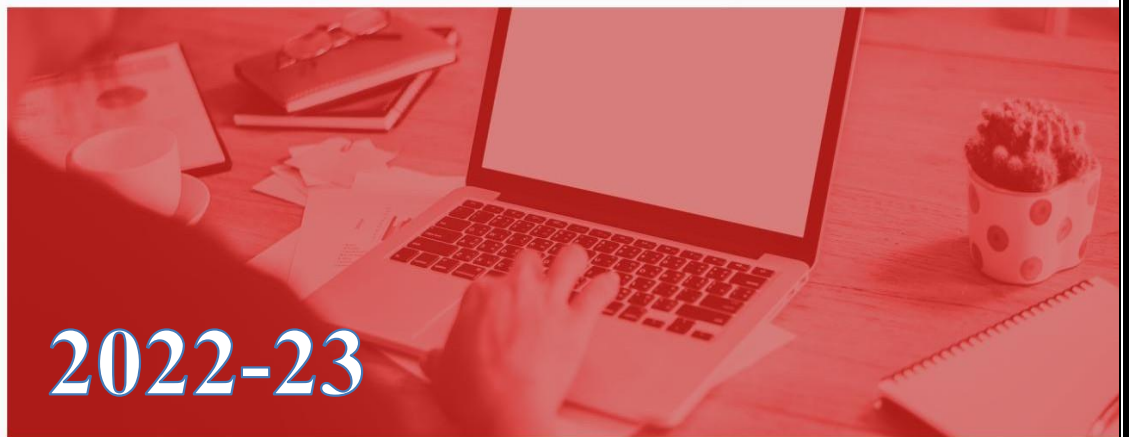Department of Computer Science and Applications

Board of Studies

B.Sc (Maths, Statistics, Data Science)

2022-23

Government College (A), Rajahmundry
Accredited with 'A+' grade by NAAC

PROCEEDINGS OF THE PRINCIPAL
GOVERNMENT AUTONOMOUS COLLEGE, RAJAHMUNDRY

PRESENT: Dr.R. David Kumar, M.Sc., M.Phil.., Ph.D.

RC. NO. 152/GCRJY/ACAD. CELL//BOS/2021/, DATED. 05.07. 2022

Sub: GCRJY-**Conduct of BoS Meetings for the Academic Year 2022-23** - Regarding
***

ORDER:

       With reference to the subject cited, the lecturers-in-Charge of all the departments are hereby informed to conduct their respective Board of Studies (BoS) meetings by the end of July 2022.

       You are also informed to intimate the date of your BoS meeting well in advance to the subject experts/University nominee/Industrial Nominee/members of BoS/Student nominee concerned to get their valuable views and suggestions in the deliberations to frame the concrete syllabi for your subjects keeping in view the objectives of the college and interest of the stake holders. The date should also be indicated to Academic Cell in advance.

       You are further suggested to utilize the academic autonomy in incorporating the additional modules in the syllabi and identify the pedagogical strategies to implement the same.

Please note that your BoS document should contain the following contents in order
  a) Proceedings of the Principal pertaining to BoS
  b) Composition of BoS
  c) Table showing the Allocation of Credits   in the following table for both theory and Lab in case of science subjects

S. No	Semester	Course Code	Title of the Course (Paper)	Max. Marks (SEE)	Marks in CIA	Hrs./week			
						L	T	P	C

L= Lecture, T= Tutorial, P= Practical, C= Credits

  d) Agenda wise Resolutions adopted in the meeting with detailed discussions
  e) Table showing Members present with signatures
  f) List of Examiners & Paper setters
  g) Syllabus for each course in the **Proforma given** (both theory & Practical in case of Science subjects) followed by model question papers (theory & practical)
  h) Unit wise Assignment questions at the end of syllabus of each course

       You are requested to submit a separate document regarding addition/deletion of specific topics from the syllabus in each course (paper) with justification, if any.

All the *new Courses/certificate courses* proposed for the calendar year 2022, Seminars/ workshops, field visits, study tours for 2022-23 should be placed before the respective Board and get them approved.

You are also requested to submit 2 hard copies & 2 soft copies (CDs) of BoS document to the Academic cell along with original bills and settle the bills after completion of the BoS meeting. You can approach the Academic Cell for necessary documents.

Most Important: You are requested to submit soft & hard copies of *Resolutions (including discussion)* separately to IQAC immediately after BoS meeting is completed.

PRINCIPAL ·
GOVERNMENT AUTONOMOUS COLLEGE
RAJAHMUNDRY

Copy to:
1. Lecturers-in-Charge of all the departments
2. File

GOVERNMENT COLLEGE AUTONOMOUSRAJAHMUNDRY
(Affiliated to Adikavi Nannaya University)
Re-Accredited by NAAC with "A+" grade

BOARD OF STUDIES 2022-23

Under Graduate Programme

FOR ALL SEMESTERS

DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

S.No	Semester	Course Code	Title of the Course (Paper)	Max Marks (SEE)	Marks in CIA	L	T	P	C
1	SEM - I	DSC-101	Problem Solving in C	50	50	3	1	-	3
2		DSC-101P	Problem Solving in C Lab	50	--	-	-	3	2
3	SEM - II	DSC-102	Python for Data science	50	50	3	1	-	3
4		DSC-102P	Python for Data science Lab	50	--	-	-	3	2
5	SEM-III	DSC-103	Introduction To Data Science With R	50	50	3	1	-	3
6		DSC-103P	R Programming Lab	50	--	-	-	3	2
7	SEM-IV	DSC-104	Big Data Technology	50	50	3	1	-	3
8		DSC-104P	Big Data Technology Lab	50	--	-	-	3	2
9		DSC-105	Data Mining and Data Analysis	50	50	3	1	-	3
10		DSC-105P	Data Mining and Data Analysis Lab	50	--	-	-	3	2

GOVERNMENT COLLEGE (AUTONOMOUS), RAJAHMUNDRY
(Accredited by NAAC with "A+" Grade)
DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS
CONSOLIDATED REPORT OF BOARD OF STUDIES FOR THE YEAR 2022-23

Chairman	Mr. Suneel Kumar Duvvuri In-charge of the Department, Department Of Computer Science & Applications Government College (A), Rajahmundry.
University Nominee	Dr.V. Persis Dept. of CSC, UCEngg. AdikaviNannaya University, Rajahmundry.
Subject Expert	Smt. E. Jyothi Kiranmayi SVD Govt. College for Women, Nidadavole
Subject Expert	Mr. R V SatyanarayanaPR GDC, Kakinada
Expert from Industry	Sri Vasanth Kumar, Lead SAS Programmer , B&P Team, EMMES Services Pvt ltd, Banglore
Members	
Smt U. Sandhya Rani	Faculty Member
Sri Devaraju Hanumanthu	Faculty Member
Sri P. Narsingarao	Faculty Member
Sri. D. SeethaRamulu	Faculty Member
Kum S.Jaya Lakshmi	Faculty Member
Sri K.Ramesh	Faculty Member
Ch. Sujatha	Faculty Member
V. Sailaja	Faculty Member
N. Priyanka	Faculty Member
M. Tejaswi	Faculty Member
B. Balaparameswari	Faculty Member
S. Lakshmi Sandhya	Student

	Government College (Autonomous) Rajahmundry	Program & Semester I B.Sc. (I Sem)			
Course Code DSC101	**TITLE OF THE COURSE** Problem Solving in C				
Teaching	Hours Allocated: 60	L	T	P	C
Pre-requisites:	Basic Mathematics	3	1	-	3

Course Objectives:

1. This course aims to provide exposure to problem-solving through programming.

2. It introduces the concepts of the C Programming language.

Course Outcomes:

On Completion of the course, the students will be able to-
CO1	Understand the evolution and functionality of a Digital Computer
CO2	Develop an algorithm for solving a given problem.
CO3	Apply 'C' language constructs like Decision control statements and Iterative statements to develop various programs.
CO4	Analyze various data structures like arrays, structures and unions, pointers and files to store and retrieve data.
CO5	Apply modular programming through functions.

Course with focus on employability / entrepreneurship / Skill Development modules

Skill Development		Employability		Entrepreneurship	

Syllabus:

UNIT I: [13 Hrs]

General Fundamentals: Introduction to computers: Block diagram of a computer, characteristics and limitations of computers, applications of computers, types of computers, computer generations.

Introduction to Algorithms and Programming Languages: Algorithm – Key features of Algorithms, Flow Charts, Programming Languages – Generations of Programming Languages

– Structured Programming Language- Design and Implementation of Correct, Efficient and Maintainable Programs.

UNIT II: [13 Hrs]

Introduction to C: Introduction – Structure of C Program – Writing the first C Program – File used in C Program – Compiling and Executing C Programs – Using Comments – Keywords – Identifiers –Basic Data Types in C – Variables – Constants – I/O Statements in C- Operators in C- Programming Examples.

Decision Control and Looping Statements: Introduction to Decision Control Statements– Conditional Branching Statements – Iterative Statements – Nested Loops – Break and Continue Statement – Goto Statement

UNIT III: [12 Hrs]

Arrays: Introduction – Declaration of Arrays – Accessing elements of the Array – Storing Values in Array– Operations on Arrays – one dimensional, two dimensional and multi dimensional arrays, character handling and strings.

UNIT IV: [12 Hrs]

Functions: Introduction – using functions – Function declaration/ prototype – Function definition – function call – return statement – Passing parameters – Scope of variables – Storage Classes – Recursive functions.

Structure, Union, and Enumerated Data Types: Introduction – Nested Structures – Arrays of Structures – Structures and Functions– Union – Arrays of Unions Variables – Unions inside Structures – Enumerated Data Types.

UNIT V: [10 Hrs]

Pointers: Understanding Computer Memory – Introduction to Pointers – declaring Pointer Variables– Pointer Expressions and Pointer Arithmetic – Null Pointers - Passing Arguments to Functions using Pointer – Pointer and Arrays – Memory Allocation in C Programs – Memory Usage – Dynamic Memory Allocation – Drawbacks of Pointers

Files: Introduction to Files – Using Files in C – Reading Data from Files – Writing Data to Files

– Detecting the End-of-file – Error Handling during File Operations – Accepting Command Line Arguments.

Text books:

1. E Balagurusamy – Programming in ANSIC – Tata McGraw-Hill publications.

2. Brain W Kernighan and Dennis M Ritchie - The 'C' Programming language" - Pearson publications.

Reference books:

1. Ashok N Kamthane: Programming with ANSI and Turbo C, Pearson Edition Publications.

2. Programming in C second edition. Pradep Day, Manas Ghosh

Web Links:

1. https://nptel.ac.in/courses/106/104/106104128/

2. https://nptel.ac.in/courses/106/105/106105171/

3. https://data-flair.training/blogs/c-tutorial/

4. https://www.programiz.com/c-programming

CO-PO Mapping:

(**1: Slight [Low];**   **2: Moderate[Medium];**   **3: Substantial[High],**   **'-' : No Correlation**)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1													
CO2													
CO3													
CO4													
CO5													

Model Blue print for the question paper setter

Blue Print				
S.No.	UNIT	Short 2 M	Essay 8 M	Weightage
1	I	2	2	20.8%
2	II	2	2	20.8%
3	III	2	2	20.8%
4	IV	1	2	18.75%
5	V	1	2	18.75%
		16	80	

GOVERNMENT COLLEGE (AUTONOMOUS), RAJAHMUNDRY

(Accredited by NAAC "A+" Grade)

DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

I B.Sc (M.S.Ds)   Semester-I

Problem solving in C

MODEL QUESTION PAPER (W.E.F 2022-2023)

Time: 2 ½ Hrs.   **Max Marks: 50 M**

SECTION - I

Answer any 5 questions   **5X2= 10M**

1. Explain Block diagram of Computer.

2. Define an Algorithm. What are the key features of an algorithm?

3. Write about go to statement with syntax and example.

4. Write a short note on Dynamic memory allocation.

5. Explain pointers in arrays.

6. How to write data from files with example?

7. Write about enumerated data types.

8. Briefly explain various types of recursions.

SECTION - II

Answer following question   **5X8 = 40M**

9. Briefly explain about generations of computers.

(OR)

10. What is a Flowchart? Explain significance with an example.

11. Explain basic data types in C?

(OR)

12. Explain about iterative statements available in C.

13. What is an Array? Explain different types of arrays with examples.

14. What is a string? Explain various string handling functions available in C.

15. Define a function. Explain the passing parameter mechanism.

16. Explain about Structure with syntax and example in detail.

17. Define and use of a pointer and write a 'C' program on swapping of two numbers using pointers.

18. Explain reading from and writing data into files.

	Government College (Autonomous) Rajahmundry				
Course Code DSC-101P	**TITLE OF THE COURSE** Problem Solving in C Lab	**Program & Semester** I B.Sc. (I Sem)			
Teaching	Hours Allocated: 40 (**Lab**)	L	T	P	C
Pre-requisites:	Basic mathematical knowledge	0	0	3	2

Objectives:

1. The purpose of this course is to introduce to students to the field of programming in C language.
2. The students will be able to enhance their analyzing and problem solving skills and use the same for writing programs in C.

List of Experiments/Syllabus:

1. Write a program to check whether the given number is Armstrong or not.
2. Write a program to find the sum of individual digits of a positive integer.
3. Write a program to generate the first n terms of the Fibonacci sequence.
4. Write a program to find both the largest and smallest number in a list of integer values
5. Write a program to demonstrate refection of parameters in swapping of two integer values using Call by Value&Call by Address
6. Write a program that uses functions to add two matrices.
7. Write a program to calculate factorial of given integer value using recursive functions
8. Write a program for multiplication of two N X N matrices.
9. Write a program to perform various string operations.
10. Write a program to search an element in a given list of values.
11. Write a program to sort a given list of integers in ascending order.
12. Write a program to calculate the salaries of all employees using Employee (ID, Name, Designation, Basic Pay, DA, HRA, Gross Salary, Deduction, Net Salary) structure.

    DA is 30 % of Basic Pay

    HRA is 15% of Basic Pay

    Deduction is 10% of (Basic Pay + DA)

Gross Salary = Basic Pay + DA+ HRA

Net Salary = Gross Salary - Deduction

13. Write a program to illustrate pointer arithmetic.

14. Write a program to read the data character by character from a file.

15. Write a program to create Book (ISBN, Title, Author, Price, Pages, Publisher)structure and store book details in a file and perform the following operations

      a.Add book details

      b.Search a book details for a given ISBN and display book details, if available

      c.Update a book details using ISBN

      d.Delete book details for a given ISBN and display list of remaining Books

Reference books:

1. E Balagurusamy – Programming in ANSIC – Tata McGraw-Hill publications.

2. YashavantKanetkar - Let Us 'C' – BPB Publications.

3. Brain W Kernighan and Dennis M Ritchie - The 'C' Programming language" - Pearson publications.

Virtual Lab Links:

1. IIIT Hyderabad: Computer Programming LAB   https://cse02-iiith.vlabs.ac.in/

	Government College (Autonomous) Rajahmundry				
Course Code DSC102	**TITLE OF THE COURSE** **Python for Data Science**	**Program & Semester** I B.Sc. (II Sem)			
Teaching	Hours Allocated: 60 (**Theory**)	L	T	P	C
Pre-requisites:	Basic programming skills	3	1	-	3

Course Objectives:

1. Identify the need for data science and solve basic problems using Python built-in data types and their methods.
2. Design an application with user-defined modules and packages using OOP concept
3. Employ efficient storage and data operations using NumPy arrays.
4. Apply powerful data manipulations using Pandas.
5. Do data pre-processing and visualization using Pandas

Course Outcomes:

On Completion of the course, the students will be able to-	
CO1	Identify the need for data science and solve basic problems using Python built-in datatypes and their methods.
CO2	Understand Object Oriented concepts in Python.
CO3	Apply Numpy methods to process the data in an array.
CO4	Summarize and Compute Descriptive Statistics using Pandas.
CO5	Apply powerful data manipulations using Pandas.
CO6	Use data pre-processing and visualization using Pandas

Course with focus on employability / entrepreneurship / Skill Development modules

Skill Development		Employability		Entrepreneurship	

Syllabus:

UNIT I: **[13 Hrs]**

Introduction to Data Science - Why Python? - Essential Python libraries - Python Introduction- Features, Identifiers, Reserved words, Indentation, Comments, Built-in Data types and their Methods: Strings, List, Tuples, Dictionary, Set - Type Conversion- Operators. Decision Making- Looping- Loop Control statement- Math and Random number functions. User defined functions - function arguments & its types.

UNIT II: **[13 Hrs]**

User defined Modules and Packages in Python- Files: File manipulations, File and Directory related methods - Python Exception Handling.

OOPs Concepts -Class and Objects, Constructors – Data hiding- Data Abstraction- Inheritance.

UNIT III: **[12 Hrs]**

NumPy Basics: Arrays and Vectorized Computation- The NumPy ndarray- Creating ndarrays- Data Types for ndarrays- Arithmetic with NumPy Arrays- Basic Indexing and Slicing - Boolean Indexing- Transposing Arrays and Swapping Axes.

Universal Functions: Fast Element-Wise Array Functions- Mathematical and Statistical Methods- Sorting- Unique and Other Set Logic.

UNIT – IV: **[12 hours]**

Introduction to pandas Data Structures: Series, Data Frame and Essential Functionality: Dropping Entries- Indexing, Selection, and Filtering- Function Application and Mapping- Sorting and Ranking.

Summarizing and Computing Descriptive Statistics- Unique Values, Value Counts, and Membership. Reading and Writing Data in Text Format.

UNIT –V: **[10 hours]**

Data Cleaning and Preparation: Handling Missing Data - Data Transformation: Removing Duplicates, Transforming Data Using a Function or Mapping, Replacing Values, Detecting and Filtering Outliers- String Manipulation: Vectorized String Functions in pandas.

Plotting with pandas: Line Plots, Bar Plots, Histograms and Density Plots, Scatter or Point Plots.

Text books:

1. Y. Daniel Liang, "Introduction to Programming using Python", Pearson, 2012.

2. Wes McKinney, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", O'Reilly, 2nd Edition, 2018.

3. Jake VanderPlas, "Python Data Science Handbook: Essential Tools for Working with Data", O'Reilly, 2017.

Reference books:

1. Wesley J. Chun, "Core Python Programming", Prentice Hall, 2006.

2. Mark Lutz, "Learning Python", O'Reilly, 4th Edition, 2009.

Web Links:

1. https://www.edx.org/course/python-basics-for-data-science

2. https://www.edx.org/course/analyzing-data-with-python

3. https://www.coursera.org/learn/python-plotting?specialization=data-science-python

4. https://www.programmer-books.com/introducing-data-science-pdf/

5. https://www.cs.uky.edu/~keen/115/Haltermanpythonbook.pdf

CO-PO Mapping:

(1: Slight [Low];   2: Moderate[Medium];   3: Substantial[High],   '-' : No Correlation)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1													
CO2													
CO3													
CO4													
CO5													

Model Blue print for the question paper setter

S.No.	UNIT	Short 2 M	Essay 8 M	Weightage
Blue Print				
1	I	2	2	20.8%
2	II	2	2	20.8%
3	III	2	2	20.8%
4	IV	1	2	18.75%
5	V	1	2	18.75%
		16	80	

GOVERNMENT COLLEGE (AUTONOMOUS), RAJAHMUNDRY

(Accredited by NAAC "A+" Grade)

DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

I B.Sc (M.S.Ds)   Semester-II

Python for Data Science

MODEL QUESTION PAPER (W.E.F 2022-2023)

Time: 2 ½ Hrs.   **Max Marks: 50 M**

SECTION - I

Answer any 5 questions   **5X2= 10M**

1. Write a short note on lists.

2. What is dictionary? Explain with example.

3. Write about constructors in python.

4. Write a short note data hiding in python.

5. Explain about indexing and slicing in python.

6. Write a short note on data types of ndarrays?

7. Write a short note on data frame.

8. Write a short note on writing data in text format.

SECTION - II

Answer following question   **5X8 = 40M**

9. Briefly explain about loop control statements in python.

(OR)

10. Explain about decision making statements in python.

11. Explain python exception handling.

(OR)

12. Explain File manipulation in Python.

13. Explain about Indexing and slicing operations.

(OR)

14. Explain about Mathematical and statistical methods in Python

.

15. Explain about Pandas Data structures.

(OR)

16. Explain about summarizing and descriptive statistics.

17. Explain about string manipulation in python.

(OR)

18. Explain about plotting with Pandas.

	Government College (Autonomous) Rajahmundry				
Course Code DSC102P	**TITLE OF THE COURSE** Python for Data Science Lab	**Program & Semester** I B.Sc. (II Sem)			
Teaching	Hours Allocated: 40 (**Lab**)	L	T	P	C
Pre-requisites:	Basic Programming skills	0	0	3	2

Objectives:

1. Implement simple programs in Python.

2. Implement programs related to various structures like arrays, lists, Data frames, etc.

3. Implement programs related to files.

4. Implement applications related to data science.

List of Experiments:

1. Perform Creation, indexing, slicing, concatenation and repetition operations on Python built-in data types: Strings, List, Tuples, Dictionary, Set

2. Apply Python built-in data types: Strings, List, Tuples, Dictionary, Set and their methods to solve any given problem.

3. Handle numerical operations using math and random number functions

4. Create user-defined functions with different types of function arguments.

5. Create packages and import modules from packages.

6. Perform File manipulations- open, close, read, write, append and copy from onefile to another.

7. Write a program for Handle Exceptions using Python Built-in Exceptions

8. Write a program to implement OOP concepts like Data hiding and Data Abstraction.

9. Create NumPy arrays from Python Data Structures, Intrinsic NumPy objects and Random Functions.

10. Manipulation of NumPy arrays- Indexing, Slicing, Reshaping, Joining and Splitting.

11. Computation on NumPy arrays using Universal Functions and Mathematical

methods.

12. Load an image file and do crop and flip operation using NumPy Indexing.

13. Create Pandas Series and Data Frame from various inputs.

14. Import any CSV file to Pandas Data Frame and perform the following:

- Visualize the first and last 10 records

- Get the shape, index and column details

- Select/Delete the records (rows)/columns based on conditions.

- Perform ranking and sorting operations.

- Do required statistical operations on the given column

15. Import any CSV file to Pandas Data Frame and perform the following:

(a) Handle missing data by detecting and dropping/ filling missing values.
(b) Transform data using apply () and map() method.
(c) Detect and filter outliers.
(d) Perform Vectorized String operations on Pandas Series.
(e) Visualize data using Line Plots, Bar Plots, Histograms, Density Plots and Scatter Plots.

Reference books:

1. Wesley J. Chun, "Core Python Programming", Prentice Hall, 2006.

Virtual Lab Links:

1. https://www.iiitmk.ac.in/DAVirtalLab/#work

	Government College (Autonomous) Rajahmundry		Program & Semester II B.Sc. (III Sem)			
Course Code DSC103	**TITLE OF THE COURSE** Introduction To Data Science With R					
Teaching	Hours Allocated: 60 (**Theory**)		L	T	P	C
Pre-requisites:	Database Management Systems		3	1	-	3

Course Objectives:

1. This course will introduce students to the collection. Preparation, analysis, modeling and visualization of data, covering both conceptual and practical issues.
2. Case studies from diverse fields will be presented, and hands-on use of statistical and data manipulation software will be included.

Course Outcomes:

On Completion of the course, the students will be able to-	
CO1	Recognize various disciplines that contribute to a successful data science effort Understand the processes of data science - identifying the problem to be solved, data collection, preparation, modeling, evaluation and visualization
CO2	Be aware of the challenges that arise in data sciences Develop and appreciate various techniques for data modeling and mining Be cognizant of ethical issues in many data science tasks
CO3	Be comfortable using commercial and open source tools such as the R language and its associated libraries for data analytics and visualization Learn skills to analyze real time problems using R
CO4	Able to use basic R data structures in loading, cleaning the data and preprocessing the data Able to do the exploratory data analysis on real time datasets
CO5	Able to understand and implement Linear Regression Able to understand and use - lists, vectors, matrices, dataframes, etc.

Course with focus on employability / entrepreneurship / Skill Development modules

Skill Development		Employability		Entrepreneurship	

Syllabus:

UNIT-I: [13 Hrs]

Introduction to Data Science- Introduction- Definition - Data Science in various fields - Examples - Impact of Data Science - Data Analytics Life Cycle - Data Science Toolkit - Data Scientist - Data Science Team

Understanding data: Introduction – Types of Data: Numeric – Categorical – Graphical – High Dimensional Data – Classification of digital Data: Structured, Semi-Structured and Un-Structured - Example Applications. Sources of Data: Time Series – Transactional Data – Biological Data – Spatial Data – Social Network Data – Data Evolution.

UNIT-II: [13 Hrs]

Introduction to R- Features of R - Environment - R Studio. Basics of R-Assignment - Modes - Operators - special numbers - Logical values - Basic Functions - R help functions - R Data Structures - Control Structures. Vectors: Definition- Declaration - Generating - Indexing - Naming - Adding & Removing elements - Operations on Vectors - Recycling - Special Operators - Vectorized if- then else-Vector Equality – Functions for vectors - Missing values - NULL values - Filtering & Subsetting.

UNIT-III: [12 Hrs]

Matrices - Creating Matrices - Adding or Removing rows/columns - Reshaping - Operations - Special functions on Matrices.

Lists - Creating List – General List Operations - Special Functions - Recursive Lists. Data Frames - Creating Data Frames - Naming - Accessing - Adding - Removing - Applying Special functions to Data Frames - Merging Data Frames- Factors and Tables.

UNIT-IV: [12 Hrs]

Input / Output – Reading and Writing datasets in various formats - Functions - Creating User-defined functions - Functions on Function Object - Scope of Variables - Accessing Global, Environment - Closures - Recursion. Exploratory Data Analysis - Data Preprocessing - Descriptive Statistics - Central Tendency - Variability - Mean - Median - Range - Variance - Summary - Handling Missing values and Outliers - Normalization

Data Visualization in R : Types of visualizations - packages for visualizations - Basic Visualizations,

Advanced Visualizations and Creating 3D plots.

UNIT-V: **[10 Hrs]**

Inferential Statistics with R - Types of Learning - Linear Regression- Simple Linear Regression - Implementation in R - functions on lm() - predict() - plotting and fitting regression line. Multiple Linear Regression - Introduction -comparison with simple linear regression - Correlation Matrix - F-Statistic - Target variables Vs Predictors - Identification of significant features - Implementation of Multiple Linear Regression in R.

Text books:

1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.
2. Jure Leskovec, Anand Rajaraman, Jeffrey D.Ullman, "Mining of Massive Datasets",Cambridge University Press, 2014.
3. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012.
4. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013.
   5.Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, Abhijit Dasgupta, "Practical

Reference books:

1. Nathan Yau, "Visualize This: The FlowingData Guide to Design, Visualization, and Statistics", Wiley, 2011.
2. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, "Professional Hadoop Solutions", Wiley, ISBN: 9788126551071, 2015.

Web Links:

1. https://nptel.ac.in/courses/111104146

CO-PO Mapping:

(1: Slight [Low];   2: Moderate[Medium];   3: Substantial[High],   '-' : No Correlation)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1													
CO2													
CO3													
CO4													
CO5													

Model Blue print for the question paper setter

Blue Print				
S.No.	UNIT	Short 2 M	Essay 8 M	Weightage
1	I	2	2	20.8%
2	II	2	2	20.8%
3	III	2	2	20.8%
4	IV	1	2	18.75%
5	V	1	2	18.75%
		16	80	

GOVERNMENT COLLEGE (AUTONOMOUS), RAJAHMUNDRY

(Accredited by NAAC "A+" Grade)

DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

I B.Sc (M.S.Ds)   Semester-III

Python for Data Science

MODEL QUESTION PAPER (W.E.F 2022-2023)

Time: 2 ½ Hrs.   **Max Marks: 50 M**

SECTION –I

Answer Any Five Questions of the following   5 X 2M=10 M

1.   Explain the properties of No-SQL

2.   How working with data from files

3.   Explain the Memorization Methods

4.   Write a short note evaluating clustering models

5.   Why is R important for data science

6.   What is ordered and unordered factors

7.   What is a block and block scanner in HDFS

8.   Explain exporting graph

SECTION - II

Answer Any Five Questions of the following   :   5 X 8M=40 M

9.   Explain about classification of digital DATA.

(OR)

10. Explain about Data analytics life cycle.

11. Explain about features of R environment.

(OR)

12. Explain about control structures in R.

13. Explain about functions in matrices.

(OR)

14. Explain about creating and merging data frames.

15. Explain about data visualization techniques in R.

(OR)

16. Explain about reading and writing data sets various formats.

17. Explain about Implementation of Multiple Linear Regressions in R.

(OR)

18. Explain about types of learning in R.

	Government College (Autonomous) Rajahmundry				
Course Code DSC103P	**TITLE OF THE COURSE** **R Programming Lab**	**Program & Semester** II B.Sc. (III Sem)			
Teaching	Hours Allocated: 40 (**Lab**)	L	T	P	C
Pre-requisites:	Python Programming	0	0	3	2

Objectives:

1. This course will introduce students to the collection. Preparation, analysis, modeling and visualization of data, covering both conceptual and practical issues

List of Experiments

1. Installing R and R studioCreate a folder DS_R and make it a working directory. Display the current working directory

2. installing the "ggplot2", "caTools", "CART" packages

3. load the packages "ggplot2", "caTools".

4. Basic operations in r

5. Working with Vectors:

   • Create a vector v1 with elements 1 to 20.

   • Add 2 to every element of the vector v1.

   • Divide every element in v1 by 5

   • Create a vector v2 with elements from 21 to 30. Now add v1 to v2.

6. Getting data into R, Basic data manipulation

7. Using the data present in the table given below, create a Matrix "**M**"

	C1	C2	C3	C4	C5
C1	0	12	13	8	20
C2	12	0	15	28	88
C3	13	15	0	6	9
C4	8	28	6	0	33
C5	20	88	9	33	0

- Find the pairs of cities with shortest distance.

8. Consider the following marks scored by the 6 students

Section	Student no	M1	M2	M3
A	1	45	54	45
A	2	34	55	55
A	3	56	66	64
B	1	43	44	45
B	2	67	76	78
B	3	76	68	37

- create a data structure for the above data and store in proper positions with proper names
- display the marks and totals for all students
- Display the highest total marks in each section.
- Add a new subject and fill it with marks for 2 sections.

9. Three people denoted by P1, P2, P3 intend to buy some rolls, buns, cakes and bread. Each of them needs these commodities in differing amounts and can buy them in two shops S1, S2. The individual prices and desired quantities of the commodities are given in the following table "demand.

	price	
	S1	S2
Roll	1.5	1
Bun	2	2.5
Cake	5	4.5
Bread	16	17

	demand.quantity			
	Roll	Bun	Cake	Bread
P1	6	5	3	1
P2	3	6	2	2
P3	3	4	3	1

- Create matrices for above information with row names and col names.
- Display the demand.quantity and price matrices
- Find the total amount to be spent by each person for their requirements in each shop
- Suggest a shop for each person to buy the products which is minimal.

10. Consider the following employee details:

employee details as follows	
emp_no:1	
name: Ram	
salary	
	basic: 10000
	hra: 2500
	da: 4000
deductions	
	pf: 1100
	tax: 200
total salary	
	gs(Gross Salary):
	ns(Net Salary)

- Create a list for the employee data and fill gross and net salary.
- Add the address to the above list
- display the employee name and address
- remove street from address
- remove address from the List.

11. Loops and functions - Find the factorial of a given number

12. Implementation of Data Frame and its corresponding operators and functions

13. Implementation of Reading data from the files and writing output back to the specified file

14. Treatment of NAs, outliers, Scaling the data, etc

15. Applying summary() to find the mean, median, standard deviation, etc

16. Implementation of Visualizations - Bar, Histogram, Box, Line, scatter plot, etc.

17. Implementation of Linear and multiple Linear Regression

18. Fitting regression line

Reference books:

1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.

Virtual Lab Links:

1. https://vlab.amrita.edu/?sub=3&brch=311&sim=1835&cnt=2

	Government College (Autonomous) Rajahmundry	Program & Semester II B.Sc. (IV Sem)			
Course Code DSC104	**TITLE OF THE COURSE Big Data Technology**				
Teaching	Hours Allocated: 60 (**Theory**)	L	T	P	C
Pre-requisites:	Basic programming skills	3	1	-	3

Course Objectives:

1. This course provides practical foundation level training that enables immediate and effective participation in big data projects.

2. The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem.

Course Outcomes:

On Completion of the course, the students will be able to-	
CO1	Learn tips and tricks for Big Data use cases and solutions. Acquire knowledge of HDFS components , Namenode, Datanode, etc.
CO2	Acquire knowledge of storing and maintaining data in cluster, reading data from and writing data to Hadoop cluster. Able to maintain files in HDFS
CO3	Able to write Map Reduce applications to access data present on HDFS Able to read different formats of files into map-reduce application
CO4	Able to develop MapReduce applications to analyze Big Data related to the real world use cases. Able to write MapReduce applications that can take data from multiple datasets and join them
CO5	Able to optimize the performance of Map-Reduce application.

Course with focus on employability / entrepreneurship / Skill Development modules

Skill Development		Employability		Entrepreneurship	

Syllabus:

UNIT-I: [13 Hrs]

Introduction to Big Data

Introduction –Distributed File System – Big Data and its importance, Characteristics of Big Data, Limitation of Conventional Data Processing Approaches, Need of big data frameworks, Big data analytics, Limitations of Big Data and Challenges, Big data applications

UNIT-II: [13 Hrs]

Hadoop: Basic Concepts of Hadoop and its features -The Hadoop Distributed File System (HDFS)- Anatomy of a Hadoop Cluster - Hadoop cluster modes - Hadoop Architecture, Hadoop Storage - Hadoop daemons (Name node-Secondary name node-Job tracker-Task tracker-Data node,etc) - Anatomy of Read & Write operations – Interacting HDFS using command-line (HDFS Shell and FS shell commands) -Interacting HDFS using Java APIs – Dataflow – Blocks –Replica - YARN.

UNIT-III: [12 Hrs]

Hadoop Ecosystem Components – Schedulers- Fair and Capacity, Hadoop 2.0 Vs Hadoop 3.0 and its new features.

Hadoop Cluster Setup – SSH & Hadoop Configuration –HDFS Administering – Monitoring & Maintenance.

UNIT-IV [12 Hrs]

Hadoop MapReduce - Introduction - Phases in MapReduce Framework - Anatomy of MapReduce Job run - Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

Understanding Basic MapReduce Pogram (WordCount program): The Driver Code - The Mapper class - The Reducer class.

UNIT-V [10 Hrs]

<mark>Writing first MapReduce Program</mark> - Hadoop's Streaming API - Using Eclipse for Rapid Development – YARN Vs MapReduce Advanced MapReduce Concepts: Partitioner – Combiner – Joins – Map-side Join – Reduce-side Join - <mark>Case Study:</mark> Weblog Analysis done using Mapper, Reducer, Combiner, Partitioner, etc.

Text books:

1. Boris lublinsky, Kevin t. Smith Alexey Yakubovich, "Professional Hadoop Solutions". Wiley, ISBN : 9788126551071, 2015.
2. Chris Eaton, Dirk Deroos et al., "Understanding Big Data", McGraw Hill , 2010.
3. Tom White, "HADOOP" : The definitive Guide", O Reilly 2012.

Reference books:

1. Srinath Perera, Thilina Gunarathne, "Hadoop MapReduce Cookbook", PACKT publishing, 2013.

Web Links:

1. https://nptel.ac.in/courses/106104189

CO-PO Mapping:

(**1: Slight [Low];** **2: Moderate[Medium];** **3: Substantial[High],** **'-' : No Correlation**)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1													
CO2													
CO3													
CO4													
CO5													

Model Blue print for the question paper setter

Blue Print				
S.No.	UNIT	Short 2 M	Essay 8 M	Weightage
1	I	2	2	20.8%
2	II	2	2	20.8%
3	III	2	2	20.8%
4	IV	1	2	18.75%
5	V	1	2	18.75%
		16	80	

GOVERNMENT COLLEGE (AUTONOMOUS), RAJAHMUNDRY

(Accredited by NAAC "A+" Grade)

DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

I B.Sc (M.S.Ds)   Semester-IV

Big Data Technology

MODEL QUESTION PAPER (W.E.F 2022-2023)

Time: 2 ½ Hrs.   **Max Marks: 50 M**

SECTION –I

Answer Any Five Questions of the following   **5 X 2M=10 M**

1. Write a short note on Big data applications.

2. Write a short note on need of Big Data frame work.

3. Write a short note on Hadoop features.

4. Write a short note Read and Write operations.

5. Write a short note on Hadoop 2.0 vs Hadoop 3.0.

6. Write a short note on hadoop configurations.

7. Write a short note on MapReduce types.

8. Write a short note on Map side Join.

SECTION - II

Answer Any Five Questions of the following   :   **5 X 8M=40 M**

9. Explain characteristics and limitations of Big data.

(OR)

10. Explain about Limitations of conventional data processing approaches.

11. Explain about how to Interact HDFS using Java APIs.

(OR)

12. Explain about Hadoop architecture.

13. Explain about HDFS Monitoring and Maintenance.

(OR)

14. Explain about HDFS Schedulers.

15. Explain about phases in Map reduce frame work.

(OR)

16. Explain about Map reduce features.

17. Explain about Weblog analysis using Mapper or Reducer.

(OR)

18. Explain about Hadoop streaming API.

	Government College (Autonomous) Rajahmundry				
Course Code CSC165P	**TITLE OF THE COURSE** Big Data Technology LAB	**Program & Semester** II B.Sc. (IV Sem)			
Teaching	Hours Allocated: 40 (**Lab**)	L	T	P	C
Pre-requisites:	Basic Programming skills	0	0	3	2

1. This course provides practical foundation level training that enables immediate and effective participation in big data projects.

2. The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem

List of Experiments:

1. Implement the following Data Structures in Java

   a) Linked Lists

   b) Stacks

   c) Queues

   d) Set

   e) Map

2. Hadoop Cluster Setup

   (i)   Perform setting up and Installing Hadoop in its three operating

          modes: Standalone

          Pseudo

          distributed

          Fully

          distributed

   (ii)   Use web based tools to monitor your Hadoop setup.

3. Implement the following file management tasks in Hadoop:

- o Adding files and directories, List the files and directories
- o Retrieving files
- o Deleting files
- o Copying files from one folder to another in HDFS
- o Copying files from Local File System to HDFS

4. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm

5. Write a Map Reduce program that mines weather data (NCDC). Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented. Data available at: ftp://ftp.ncdc.noaa.gov/pub/data/noaa/.

- Find average, max and min temperature for each year in NCDC data set
- Filter the readings of a set based on value of the measurement, Output the line of input files associated with a temperature value greater than 30.0 and store it in a separate file.

6. Implement Matrix Multiplication program with Hadoop Map Reduce.

7. Stop word elimination problem:

   Input:

- A large textual file containing one sentence per line
- A small file containing a set of stop words (One stop word per line)

   Output:

- A textual file containing the same sentences of the large input file without the words appearing in the small file.

8. Write a MapReduce Application to implement Combiners

9. Write a MapReduce Application to implement Reduce-side Join

10. Write a MapReduce Application to implement Map-side Join

Reference books:

1. Srinath Perera, Thilina Gunarathne, "Hadoop MapReduce Cookbook", PACKT

publishing, 2013.

Virtual Lab Links:

1. https://futureskillsprime.in/courses/big-data-hadoop-first-virtual-lab

	Government College (Autonomous) Rajahmundry				
Course Code DSC105	**TITLE OF THE COURSE** Data Mining and Data Analysis	**Program & Semester** II B.Sc. (IV Sem)			
Teaching	Hours Allocated: 60 (**Theory**)	L	T	P	C
Pre-requisites:	Basic programming skills	3	1	-	3

Course Objectives:

1. To learn data analysis techniques.
2. To understand Data mining techniques and algorithms.
3. Comprehend the data mining environments and application.

Course Outcomes:

On Completion of the course, the students will be able to-	
CO1	To understand and demonstrate data mining Compare various conceptions of data mining as evidenced in both research and application.
CO2	Characterize various kinds of patterns that can be discovered by association rule mining.
CO3	Evaluate mathematical methods underlying the effective application of data mining
CO4	To Analyze the data using statistical methods
CO5	Gain hands-on skills and experience on data mining tools.

Course with focus on employability / entrepreneurship / Skill Development modules

Skill Development		Employability		Entrepreneurship	

Syllabus:

UNIT-I [13 Hrs]

Data mining - KDD Vs Data Mining, Stages of the Data Mining Process-Task Primitives, Data Mining Techniques – Data Mining Knowledge Representation. Major Issues in Data Mining – Measurement and Data – Data Preprocessing – Data Cleaning - Data transformation- Feature

Selection - Dimensionality reduction

UNIT-II   **[13 Hrs]**

Predictive Analytics

Classification and Prediction - Basic Concepts of Classification and Prediction, General Approach to solving a classification problem- Logistic Regression - LDA - Decision Trees: Tree Construction Principle – Feature Selection measure – Tree Pruning - Decision Tree construction Algorithm, Random Forest, Bayesian Classification-Accuracy and Error Measures- Evaluating the Accuracy of the classifier / predictor- Ensemble methods and Model selection.

UNIT-III   **[12 Hrs]**

Classification and Descriptive Analytics

Rule Based Classification – Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction.

Descriptive Analytics - Mining Frequent Item sets - Market based model – Association and Sequential Rule Mining

UNIT - IV   **[12 Hrs]**

Cluster Analysis

Cluster Analysis: Basic concepts and Methods – Cluster Analysis – Partitioning methods – Hierarchical methods – Density Based Methods – Grid Based Methods – Evaluation of Clustering – Advanced Cluster Analysis: Probabilistic model based clustering – Clustering High – Dimensional Data – Clustering Graph and Network Data – Clustering with Constraints- Outlier Analysis.

UNIT-V   **[10 Hrs]**

Factor Analysis

Factor Analysis: Meaning, objectives and Assumptions, Designing a factor analysis, Deriving factors and assessing overall factors, Interpreting the factors and validation of factor analysis.

Text books:

1. Adelchi Azzalini, Bruno Scapa, "Data Analysis and Data mining" , 2nd Ediiton, Oxford Univeristy Press Inc., 2012.

2. Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", 3rd Edition, Morgan Kaufmann Publishers, 2011.

3. Alex Berson and Stephen J. Smith, "Data Warehousing, Data Mining & OLAP", 10th Edition, TataMc Graw Hill Edition , 2007.

4. G.K. Gupta, "Introduction to Data Mining with Case Studies", 1st Edition, Easter Economy Edition, PHI, 2006.

Reference books:

1. Joseph F Hair, William C Black etal, "Multivariate Data Analysis", Pearson Education, 7th edition, 2013.

Web Links:

1. https://nptel.ac.in/courses/106105174

CO-PO Mapping:

(1: Slight [Low];   2: Moderate[Medium];   3: Substantial[High],   '-' : No Correlation)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PSO1	PSO2	PSO3
CO1													
CO2													
CO3													
CO4													
CO5													

Model Blue print for the question paper setter

Blue Print				
S.No.	UNIT	Short 2 M	Essay 8 M	Weightage
1	I	2	2	20.8%
2	II	2	2	20.8%
3	III	2	2	20.8%
4	IV	1	2	18.75%
5	V	1	2	18.75%
		16	80	

GOVERNMENT COLLEGE (AUTONOMOUS), RAJAHMUNDRY

(Accredited by NAAC "A+" Grade)

DEPARTMENT OF COMPUTER SCIENCE & APPLICATIONS

I B.Sc (M.S.Ds)   Semester-IV

Data Mining and Data Analysis

MODEL QUESTION PAPER (W.E.F 2022-2023)

Time: 2 ½ Hrs.   **Max Marks: 50 M**

SECTION –I

Answer Any Five Questions of the following   **5 X 2M=10 M**

1. What is data mining and what are the stages in data mining.

2. Write a short note on Data mining task primitives.

3. Write a short note on Tree pruning.

4. Write a short note classification vs. prediction.

5. Define the terms support and confidence.

6. Write a short note on rule based classification.

7. Write a short note on partition based clustering technique.

8. What are the objectives of factor analysis?

SECTION - II

Answer Any Five Questions of the following   :   **5 X 8M=40 M**

9. Explain about various data preprocessing techniques.

(OR)

10. Explain about design issues in data mining.

11. Explain about Decision tree induction algorithm.

(OR)

12. Explain about Bayesian Classification.

13. Explain about Support Vector Machines.

(OR)

14. Explain about Mining frequent Item sets using FP tree algorithm.

15. Explain about Grid based methods.

(OR)

16. Explain about Hierarchical methods.

17. Explain about assessing overall factors.

(OR)

18. Explain about Interpreting and validation of factor analysis.

	Government College (Autonomous) Rajahmundry	Program & Semester			
Course Code CDS-105P	TITLE OF THE COURSE DATA ACQUISITION AND ANALYSIS LAB	II B.Sc. (IV Sem)			
Teaching	Hours Allocated: 30 (Lab)	L	T	P	C
Pre-requisites:	Basic Knowledge of Programming	0	0	3	2

Course Objectives:

1. This course provides practical foundation level training that enables immediate and effective participation in data analysis techniques.
2. To understand Data mining techniques and algorithms. Comprehend the data mining environments and application

List of Experiments/Syllabus:

1. Data Analysis – Getting to know the Data (Using ORANGE WEKA or R Programming)
   - Parametric – Means, T-Test, Correlation
   - Prediction for numerical outcomes – Linear regression, Multiple Linear Regression
   - Correlation analysis
   - Preparing data for analysis
     - o Pre-Processing techniques

2. Data Mining (Using ORANGE WEKA or R Programming)
   - Implement clustering algorithm
   - Implement Association Rule mining
   - Implement classification using
     - o Decision tree
     - o Back Propagation

- - *Logistic Regression*
  - *Decision Tree*
  - *Random Forest*
  - *Naive Bayes*
  - *Support Vector Machines*
- Visualization methods