SRI VENKATESWARA UNIVERSITY
BCA (Big Data) Honours– W.E.F. 2024-25

Year	Semester	Course	Titleofthe Course	No. of Hrs/Week	No. of Credits
II	IV	9	Python Programming	3	3
			Python ProgrammingLab	2	1
		10	Big Data Technology	3	3
			Big Data Technology Lab	2	1
		11	Statistical Techniques	3	3
			Statistical Techniques Lab	2	1

Note:
Course-9 Python Programming is common to all BCA Programmes General/ Artificial Intelligence/ Big Data / Data Science / Cloud Computing Specializations.

Course-11 Statistical Techniques is common to BCA Programmes Artificial Intelligence/ Big Data/ Data Science / Cloud Computing Specializations.

SRI VENKATESWARA UNIVERSITY::TIRUPATI

BCA General/Artificial Intelligence/Big Data/Cloud Computing/ Data Science

Honours

II Year IV Semester

COURSE9:PYTHONPROGRAMMING

(w.e.f. 2024-25)

Theory **Credits: 3** **3 hrs/week**

Course Objectives:

Python is a language with a simple syntax, and a powerful set of libraries. It is an interpreted language, with a rich programming environment, including a robust debugger and profiler. While it is easy for beginners to learn, it is widely used in many scientific areas for data exploration.

Course Outcomes:

Upon successful completion of the course, a student will be able to:

1. Examine Python syntax and semantics and be fluent in the use of Python flow control and functions.
2. Demonstrate proficiency in handling Strings and File Systems.
3. Create, run and manipulate Python Programs using core data structures like Lists, Dictionaries and use Regular Expressions.
4. Interpret the concepts of Object-Oriented Programming as used in Python.

Unit-I

Getting Started with Python: Introduction to Python, Python Keywords, Identifiers, Variables, Comments, Data Types, Operators, Input and Output, Type Conversion, Debugging. Flow of Control, Selection, Indentation, Repetition, Break and Continue Statement, Nested Loops.

Strings-String Operations, Traversing a String, String handling Functions.

Case Study:

1. Study the features that make Python different from Procedural Languages.

Unit-II

Functions: Functions, Built-in Functions, User Defined Functions, recursive functions, Scope of a Variable

Python and OOP: Defining Classes, Defining and calling functions passing arguments, Inheritance, polymorphism, Modules– date time, math, Packages.

Exception Handling- Exception in python, Types of Exception, User-defined Exceptions.

Case Study:

1. Present a report of how Exception handling is different from JAVA Exceptional Handling.

Unit-III

List: Introduction to List, List Operations, Traversing a List, List Methods and Built-in Functions.

Tuples and Dictionaries: Introduction to Tuples, Tuple Operations, Tuple Methods and Built-in Functions, Nested Tuples. Introduction to Dictionaries, Dictionaries are Mutable, Dictionary Operations, Traversing a Dictionary, Dictionary Methods and Built-in functions.

Case Study:

1. What are the special features of dictionaries and try to analyze about the same features in any other language.

Unit-IV

Introduction to NumPy : Array, NumPyArray, Indexing and Slicing, Operations on Arrays, Concatenating Arrays, Reshaping Arrays, Splitting Arrays, Statistical Operations on Arrays.

Data Handling: Introduction to Python Libraries, Series, Data Frame, Importing and Exporting Data between CSV Files and Data Frames.

CaseStudy:

1. Present a paper on advanced features of NumPy.

Unit-V

Plotting Data using Matplotlib: Introduction, Plotting using Matplotlib – Line chart, Bar chart, Histogram, Scatter Chart, Pie Chart.

Database Connectivity: Importing MySQL for Python, Connecting with a database, Forming a query in MySQL, Passing a query to MySQL.

CaseStudy:

1. Present a paper on the features and advantages of MySQL compared to other commercial Databases.

References:

1. Mark Lutz, Learning Python,5th Ed. O‟REILLY
2. Core Python Programming by Dr. R. Nageswara Rao
3. Problem Solving and Python Programming by E.Balaguru Swamy
4. Python programming: using problem solving approach by Reema Thareja.
5. Albert Lukaszewski, MySQL for Python, Packet Publishing

BCA General/Artificial Intelligence/Big Data/Cloud Computing/ Data Science Honours

II Year IV Semester

COURSE9:PYTHONPROGRAMMING

Practicals **Credits: 1** **2hrs/week**

Lab Programs

1. Write a Program to check whether given number is Armstrong or not.
2. Write a Program to check whether given number is perfect or not.
3. Write a program to find factorial of given number using recursive function
4. Write a program to implement inheritance and polymorphism
5. Demonstrate a python code to print try, except and finally block statements
6. Write a program to demonstrate String handling functions
7. Write a program to input n numbers from the user. Store these numbers in a tuple. Print the maximum and minimum number from this tuple.
8. Write a program to enter names of employees and their salaries as input and store them in a dictionary
9. Write a program to implement statistical operations on arrays using numPy
10. Write a program to import and export CSV file to DataFrame.
11. Create the DataFrame Sales containing yearwise sales and perform basic operation on it.
12. Visualize the plots using matplot lib.
13. Write a program to connect with MySQL database and perform CRUD(Create, Read, Update and Delete) operations

MODEL QUESTION PAPER

SRI VENKATESWARA UNIVERSITY::TIRUPATI

BCA General/Artificial Intelligence/Big Data/Cloud Computing/ Data Science

Honours

II Year IV Semester

COURSE9:PYTHONPROGRAMMING

(w.e.f. 2024-25)

Time : 3 Hrs **Max Marks : 75**

SECTION - A

Answer any Five of the following **5 X 3= 15 Marks**

1. Short answer question from Unit-1
2. Short answer question from Unit-1
3. Short answer question from Unit-2
4. Short answer question from Unit-2
5. Short answer question from Unit-3
6. Short answer question from Unit-3
7. Short answer question from Unit-4
8. Short answer question from Unit-4
9. Short answer question from Unit-5
10. Short answer question from Unit-5

SECTION - B

Answer any Five of the following **5 X 12= 60 Marks**

11. Long answer question from Unit-1
12. Long answer question from Unit-1
13. Long answer question from Unit-2
14. Long answer question from Unit-2
15. Long answer question from Unit-3
16. Long answer question from Unit-3
17. Long answer question from Unit-4
18. Long answer question from Unit-4
19. Long answer question from Unit-5
20. Long answer question from Unit-5

SRI VENKATESWARA UNIVERSITY::TIRUPATI

BCA Big DataHonours

II Year IV Semester

COURSE 10:Big Data Technology

(w.e.f. 2024-25)

Theory **Credits: 3** **3 hrs/week**

CourseObjectives:

This course provides practical foundation level training that enables immediate and effective participation in big data projects. The course provides grounding in basic and advanced methods to big data technology and tools, including MapReduce and Hadoop and its ecosystem.

Course Out comes : The students will be able to:

- Acquire knowledge of HDFS components ,Namenode, Datanode, etc.
- Acquire knowledge of storing and maintaining data in cluster, reading data from and writing data to Hadoop cluster and to maintain files in HDFS
- Able to write MapReduce applications to access data present on HDFS
- Able to read different formats of files into map-reduce application.
- Able to develop MapReduce applications to analyze Big Data related to the real world use cases.

Unit I Introduction to Big Data

Introduction -Distributed File System - BigData and itsimportance, Characteristics of Big Data, Limitationof ConventionalData Processing Approaches, Needof bigdata frameworks, Bigdataanalytics, Limitations of Big Dataand Challenges, Bigdataapplications

Unit-II Hadoop

BasicConceptsofHadoopanditsfeatures-TheHadoopDistributedFileSystem HDFS)-AnatomyofaHadoopCluster- Hadoopclustermodes-HadoopArchitecture, HadoopStorage-Hadoopdaemons(Name node-Secondarynamenode-Jobtracker-Task tracker-Datanode,etc)-Anatomyof Read&Writeoperations–InteractingHDFSusing command-line(HDFSShell andFSshellcommands)-InteractingHDFSusingJavaAPIs– Dataflow–Blocks–Replica-YARN.

Unit-III HadoopEcosystemComponents

Schedulers-FairandCapacity,Hadoop2.0 VsHadoop 3.0andits newfeatures.

Hadoop ClusterSetup –SSH&Hadoop Configuration –HDFSAdministering–Monitoring& Maintenance.

Unit-IV Hadoop MapReduce

Introduction -Phases in MapReduceFramework - Anatomy of MapReduceJobrun-Failures,JobScheduling,ShuffleandSort,TaskExecution,MapReduce Types and Formats, Map Reduce Features. Understanding BasicMapReduceProgram (WordCountprogram): TheDriverCode- TheMapperclass-TheReducer class.

Unit-V

WritingfirstMapReduceProgram-Hadoop's StreamingAPI-UsingEclipseforRapid Development –YARN Vs MapReduceAdvanced MapReduceConcepts: Partitioner–Combiner–Joins–Map-sideJoin–Reduce-sideJoin-CaseStudy:WeblogAnalysisdone usingMapper,Reducer,Combiner,Partitioner,etc.

Text References

1. Borislublinsky,Kevint. SmithAlexeyYakubovich,"ProfessionalHadoopSolutions". Wiley,ISBN:9788126551071,2015.
2. Chris Eaton,DirkDeroosetal.,"UnderstandingBigData",McGrawHill, 2010.
3. TomWhite,"HADOOP":ThedefinitiveGuide",OReilly2012.
4. SrinathPerera, ThilinaGunarathne,"HadoopMapReduceCookbook", PACKT publishing,2013.

BCA Big DataHonours

II Year IV Semester

COURSE 10:Big Data Technology

(w.e.f. 2024-25)

Practical **Credits: 1** **2hrs/week**

ListofExperiments

1. ImplementthefollowingDataStructures inJava
   a)Linked Lists
   b)Stacks
   c)Queues
   d)Set
   e)Map
2. HadoopClusterSetup
   (i) PerformsettingupandInstallingHadoopinits three operatingmodes:
   Standalone,
   Pseudo distributed,
   Fully distributed
   (ii)Usewebbasedtools tomonitoryourHadoopsetup.
3. Implementthefollowingfilemanagementtasks inHadoop:
   - oAddingfiles anddirectories,Listthefiles anddirectories
   - oRetrievingfiles
   - oDeletingfiles
   - oCopyingfilesfromonefoldertoanotherinHDFS
   - oCopyingfilesfromLocalFileSystemtoHDFS
4. RunabasicWordCountMapReduceprogramtounderstandMapReduceParadigm
5. WriteaMapReduceprogramthatmines weatherdata(NCDC).Weathersensors collectingdataeveryhouratmanylocations acrosstheglobegatheralargevolumeof logdata,whichis a goodcandidateforanalysis withMapReduce,sinceitis semi structuredandrecord-oriented.Dataavailableat: ftp://ftp.ncdc.noaa.gov/pub/data/noaa/.
   - •Findaverage,maxandmintemperatureforeachyearinNCDC dataset
   - •Filterthe readings ofasetbasedonvalueofthemeasurement,Outputthelineof inputfilesassociatedwithatemperaturevaluegreaterthan30.0andstoreitin aseparatefile.
6. ImplementMatrixMultiplicationprogramwithHadoopMapReduce.
7. Stopwordeliminationproblem:
   Input:
   - •Alargetextualfilecontainingonesentenceperline
   - •Asmallfilecontainingasetofstopwords (Onestopwordperline)

   Output:
   - •Atextualfilecontainingthesamesentencesofthelargeinputfilewithout thewords appearinginthesmallfile.
8. WriteaMapReduceApplicationtoimplementCombiners
9. WriteaMapReduceApplicationtoimplementReduce-sideJoin

10. WriteaMapReduceApplicationtoimplementMap-sideJoin

MODEL QUESTION PAPER
SRI VENKATESWARA UNIVERSITY::TIRUPATI
BCA Big DataHonours
II Year IV Semester
COURSE 10:Big Data Technology
(w.e.f. 2024-25)

Time : 3 Hrs **Max Marks : 75**

SECTION - A

Answer any Five of the following **5 X 3= 15 Marks**

1. Short answer question from Unit-1
2. Short answer question from Unit-1
3. Short answer question from Unit-2
4. Short answer question from Unit-2
5. Short answer question from Unit-3
6. Short answer question from Unit-3
7. Short answer question from Unit-4
8. Short answer question from Unit-4
9. Short answer question from Unit-5
10. Short answer question from Unit-5

SECTION - B

Answer any Five of the following **5 X 12= 60 Marks**

11. Long answer question from Unit-1
12. Long answer question from Unit-1
13. Long answer question from Unit-2
14. Long answer question from Unit-2
15. Long answer question from Unit-3
16. Long answer question from Unit-3
17. Long answer question from Unit-4
18. Long answer question from Unit-4
19. Long answer question from Unit-5
20. Long answer question from Unit-5

Note: **The question paper setter is requested to set question paper based on a model question paper and ensure coverage across all units equally**.

SRI VENKATESWARA UNIVERSITY::TIRUPATI

Common to BCA

Artificial Intelligence/Big Data/Cloud Computing/Data Science

II Year IV Semester

COURSE 11: Statistical Techniques

(w.e.f. 2024-25)

Theory **Credits: 3** **3 hrs/week**

Learning Outcomes: After successful completion of this course, the student will be able to;

1. Knowledge of Statistics and its scope and importance in various areas such as Medical, Engineering, Agricultural and Social Sciences etc.
2. Knowledge of various types of data in diagrammatic representation.
3. Brief analysis of different types of data and tabulated.
4. Knowledge of various types of data, their organization and evaluation of summary measures such as measures of central tendency and dispersion etc.
5. Insights into preliminary exploration of different types of data.

UNIT – I

Measures of Dispersion: Introduction - methods of measuring Dispersion-Range, Quartile deviation, Mean deviation from Mean, Mean deviation from Median, Calculation for individual observation, calculation on Discrete Series and Continuous Series. (Problems Only)

UNIT- II

Variance - Standard Deviation: Meaning ofVariance - Standard Deviation - calculation for Individual observation, calculation on discrete and continuous Series - Coefficient of Variation. (Problems Only)

UNIT- III

Asymmetrical Distribution(Skewness): Meaning of Symmetrical and Asymmetrical distribution – Concept of Skewness - Karl Pearson's coefficient of skewness, Bowley's Coefficient of skewness Calculation for continuous series. (Problems Only)

(Problems Only)

UNIT- IV

Measures of Correlation: Meaning of Correlation – Measure of Correlation in individual observation only – Probable Error – Karl pearson's– Spearman Rank Correlation Co-efficient. (Problems Only)

UNIT- V

Index Numbers: Unweighted Index Numbers – Simple Aggregative Method – Simple Average of Relatives method using Arithmetic Mean – Weighted Index Numbers – Laspeyre's Index Numbers, Paasche's Index Number, Fisher's Ideal Index Numbers – Time Reversal Test and Factor Reversal Test. (Problems Only)

Note :

1. Concentration on numerical problems only.

2. Proofs of theorems and Derivations of expressions are omitted.

Text Books:

1. Statistical methods - S.P. Gupta.
2. Fundamentals of Mathematical Statistics - SC Gupta and V.K. Kapoor

Reference Books:

Quantitative Techniques –Sultan Chand Publication

Suggested Co-Curricular Activities:

1. Assignments (in writing and doing forms on the aspects of syllabus content and outside the syllabus content. Shall be individual and challenging)
2. Student seminars (on topics of the syllabus and related aspects (individual activity))
   a. Quiz (on topics where the content can be compiled by smaller aspects and data (Individuals or groups as teams))
   b. Study projects (by very small groups of students on selected local real-time problems pertaining to syllabus or related areas. The individual participation and contribution of students shall be ensured (team activity))

Common to BCA
Artificial Intelligence/Big Data/Cloud Computing/Data Science
II Year IV Semester
COURSE 11: Statistical Techniques

Practicals **Credits: 1** **2 hrs/week**

Statistical Foundation - Practical (Programme implementation may be either Python/ Java/C Programming Languages)

1. Quartile Deviation
2. Mean Deviation
3. Standard Deviation.
4. Karl Pearson's Coefficient of Skewness.
5. Bowley's Coefficient of Skewness.
6. Karl Pearson's Coefficient of Correlation
7. Spearman Rank Correlation Coefficient
8. Unweighted Index Number
9. Weighted index Number
10. Time Reversal Test

MODEL QUESTION PAPER
SRI VENKATESWARA UNIVERSITY::TIRUPATI
Common to BCA
Artificial Intelligence/Big Data/Cloud Computing/Data Science
II Year IV Semester
COURSE 11: Statistical Techniques
(w.e.f. 2024-25)

Time : 3 Hrs **Max Marks : 75**

SECTION - A

Answer any Five of the following **5 X 3= 15 Marks**

NOTE:Statistical tables and Electronic Calculators are allowed

1. Calculate range and its Co-efficient

Marks	80	90	70	60	40	30

2. Calculate Co-efficient of Quartile Deviation
   $Q_1 = 40$ $Q_3 = 60$
3. What is Symmetrical and Asymmetrical Distribution
4. Calculate Karl Pearsons Co-efficient of Skewness
   Mean = 40; Median = 50; SD = 20
5. Calculate Variance

Values	10	20	30	40	50

6. Calculate Co-efficient of variation
   Mean = 40; SD = 30
7. Calculate rank correlation co-efficient

Marks in Maths	40	60	70	50	30
Marks in Satistics	80	50	40	70	40

8. Define Probable Error
9. Time Reversal Test
10. Fishers Ideal Index

SECTION - B

Answer any Five of the following **5 X 12= 60 Marks**

11. Calculate co-efficient of Quartile Deviation

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	10	20	10	30	20	10	20

12. Calculate Mean Deviation from Mean

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of	4	6	10	20	10	6	4

Students							

13. Calculate Karl Pearsons co-efficient of Skewness

Class	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	6	5	7	14	8	5	5

14. Calculate Bowleys co-efficient of Skewness

Marks	0-20	20-40	40-60	60-80	80-100	100-120	120-140
No. of Students	14	13	12	15	17	19	10

15. From the following find out who is the best and more consistent

X	40	50	60	70	80	90
Y	35	85	75	65	45	45

16. Find the co-efficient of Standard Deviation

Class	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	4	6	10	20	10	10

17. Find Karl pearsons co-efficient of correlation

X	9	8	7	6	5	4	3	2	6	7
Y	10	12	14	16	20	13	12	11	5	4

18. Calculate Spearmen's Rank correlation co-efficient

A	50	60	80	70	30	90
B	100	90	70	60	80	40

19. Calculate Fisher's Ideal Index

Commodities	2023		2022	
	Price(₹)	Quantity	Price(₹)	Quantity
A	4	20	3	10
B	5	10	2	12
C	3	30	3	5
D	4	20	2	6

20. Calculate Index Numbers Under (a) Simple Aggregative method (b) Simple Average of Relatives using AM

Commodities	Price(₹)	
	2023	2024
A	12	18
B	14	22
C	16	20