Jagat Guru Nanak Dev Punjab State Open University, Patiala

(Established by the Act No.19 of 2019 of the Legislature of the State of Punjab)

School of Sciences and Emerging Technologies

S. No	Course Name	PPR and Syllabi	Page No.
1	Bachelor of Science (Data Science)	Programme Project Report and Syllabi of Courses for Sessions (2021-22, 2022-23, 2023-24)	5-86

PROGRAMME PROJECT REPORT
Bachelor of Science (Data Science)

Content

PART-I
PROGRAMME PROJECT REPORT

A. Introduction

The school of Emerging Sciences and Technologies of Jagat Guru Nanak Dev Punjab State Open University, Patiala has started Bachelor of Science (Data Science) from the session 2020-21. The university has adopted choice-based credit system which has been introduced by University Grant Commission (UGC). This is a broad-based programme covering disciplinary, interdisciplinary and skill-based subjects. It provides flexibility for the learners to choose the subjects of their interest and also has the opportunity to grow with present digital era. The Bachelor of Science (Data Science) has 132 credits consisting of 4 categories namely Core Courses, Discipline Specific Electives, Ability Enhancement Compulsory Course and Skill Enhancement Courses. This 3-year course is of 6 semesters and each semester consists of 4 papers.

B. Mission & Objectives
▪ Mission Statement

To educate students with high quality theoretical, practical, ethical, technological and skill-oriented education in the area of Data Science and other related disciplines that can help them in their professional career and also enable them to become successful entrepreneurs. The programme's mission is to help the learners enhance their proficiency.

▪ Objectives

The programme has been framed to achieve the following objectives

- To develop an understanding and knowledge of the basic theory of Computer Science and Data Science with good foundation on theory, systems and applications.

- To be a foundation graduate Programme that will act as a feeder course for higher studies in the area of Data Science/ Machine Learning

- To acquire necessary and state-of-the-art skills to take up industry challenges in Data Science.

- The ability to synthesize the acquired knowledge, understanding and experience for a better and improved comprehension of the real-life problems

- To learn skills and tools like Machine Learning, Data mining, mathematics, statistics, to find the solution, interpret the results and make predictions for future developments.

C. Relevance of The Programme

In this age of Computers and everything being digitalized, knowledge about machines is very important. Presently Data scientist is one of the fastest-growing job titles in the industry. To cater the need of the industry a course in Data Science will prove to be very helpful for students and professionals those want to take up the job in the field of Data Sceicne. A 3-year degree course in Data Science will provide foundation skills and information not only about Data Science but also in the related fields, This will give a strong foundation to the learner. One also gets to learn programming languages used in Data Sceicne along with information about various tools and approaches used in Data Science. The bachelor Course in Data Science of Jagat Guru Nanak Dev Punjab State Open University has been designed to supply trained manpower in ever-growing IT and IT Enabled industry

D. Prospective Target Group

Passed 10+2 in Non-Medical/Commerce/Arts with Mathematics/Statistics/Computer Science/Economics as one of the optional subject or its equivalent examination stream conducted by Board recognized or established by Central/State Government through a legislation OR

Those candidates who have passed their Matriculation examination AND have also passed three year Diploma in any Trade from Punjab State Board of Technical Education & Industrial Training, Chandigarh or such Examination from any other recognized State Board of Technical Education.

Learners with above said eligibility may join this course to improve their knowledge, skills, employability, and entrepreneurship ability. The working persons and who cannot study through regular mode can continue their education through this open learning mode.

E. Appropriateness of the Programme

 The course will provide academic continuity to the learning community and will facilitate continuous professional development for the employees and entrepreneurs across the country and Punjab state, in particular. The course aims to reach the learners who are distant and those lacking access. To reach the unreached, the courses' instructions and specially prepared study material in the form of printed notes and audio-video lessons to the learners will be delivered at their doorsteps through postal correspondence and digital media like e-mail, website etc. Limited face-to-face contact sessions will be held at the study centres set up by the university as close as possible to the learner's home. Communication with the university and interaction between the teacher and the learners will be further facilitated using electronic media options like telephone, e-mails, chat sessions, videoconferencing and teleconferencing, if and when required. All of these characteristics will help learners to engage in relevant, purposeful and interesting lessons.
 Apart from this, the students will have the advantage to study at their own pace and convenience as the course can be completed in the time span ranging from six months to two years.

F. Instructional Design

The B.Sc (Data Science) is a 132 credit programme consisting of the following four categories mentioned in Table I . The programme can be completed by earning the minimum required credits under each category in a minimum 3 years period or maximum 6 years period. It is designed to provide the learners with the information and skills necessary to understand and analyse their world by introducing them to the main themes and topics of disciplines in Computer Science and Data Science. The Programme Code of this BSc (Data Science) is BSD.

Apart from this, the programme includes three non-credit yet compulsory qualifying courses: Teachings of Sikh Gurus, Human Rights, Drug abuse: Problem, prevention and Management in the first three semesters respectively. Teachings of Sikh Gurus will be taught in the first semester to familiarize the learners with the Sikh theology, rich culture and ethos of Punjab. In order to bring awareness in the learners about human rights and to sensitise them with the social issues like drug abuse, the learners will be introduced to the subjects titled Human Rights and Duties; and Drug Abuse: Problem, Prevention and Management in the subsequent semesters i.e., second and third semester, as per the official guidelines issued.

Table I: Basic Structure, Distribution of Courses, Papers and Credits

Sr No.	Course Name	No. of Papers	Credit	Total Credit
1	Core Course	12	4	48
2	Core Course Practical	12	2	24
3	Elective Course	6	4	24
4	Elective Course Practical	6	2	12
5	Ability Enhancement Compulsory Course	2	4	8
6	Skill Enhancement Elective Course	4	4	16
7	Non-Credit Qualifying Course	3	-------	----------
	Total			**132**

Note: The programme can be completed by earning the required number of credits under each category in a minimum period of 3 years (6 semesters) or in the maximum period of 6 years. The required number of credits under each category is as follows: 72 credits of Core Courses, 36 credits of Discipline Specific Electives, 8 credits of Ability Enhancement Compulsory Courses, 16 credits of Skill Enhancement courses. A credit is equivalent to 30 hours of study time comprising all learning activities (i.e. reading and comprehending the print material, listening to audios, watching videos, attending counseling sessions, teleconferencing and writing assignment responses). The programme has a mix of different types of courses in each of the six semesters. However, the total number of credits to study in each semester is 22 credits. Table 2 gives an

overview of the programme structure showing the distribution of different types of courses across the six semesters of the programme (Table-II).

Most courses of this programme are of four credits. This means that you will have to put in 120 hours (4x30) of study time to complete each of these courses. Laboratories are of 2 credit. The programme has a mix of different types of courses in each of the six semesters. However, the total number of credits to study in each semester is 22 credits. Table 2 gives an overview of the programme structure showing the distribution of different types of courses across the six semesters of the programme.

Table II: B. Sc (Data Science), Three-Year (Six Semester) CBCS Programme

Course Code	Course Title	Course Type	Credits
	Semester I		
BSDB31101T	Problem Solving using Computer	CC-1	4
BSDB31102T	Fundamental of IT	DSC-1	4
BSDB31103T	Introduction Data Science	DSC-2	4
BSDB31101P	Problem Solving using Computer Lab	CC-1	2
BSDB31102P	Fundamental of IT Lab	DSC-3	2
BSDB31103P	Introduction Data Science Lab	DSC-4	2
AE1B31104T	Effective Communication in English	AECC-1	4

TSGB31105T	Teaching of Sikh Gurus **		
	Semester II		
BSDB31201T	Operating System	CC-3	4
BSDB31202T	Statistical Foundation	DSC-5	4
BSDB31203T	Introduction to Logic	DSC-6	4
BSDB31201P	Operating System Lab	CC-3	2
BSDB31202P	Statistical Foundation Lab	DSC-7	2
BSDB31203P	Introduction to Logic	DSC-8	2
AE2B31204T	Environmental Studies	AECC-2	4
HRDB31205T	Human Rights and Duties **		0
	Semester III		
BSDB32301T	Database Management Systems	CC-3	4
BSDB32302T	Data mining and Visualization	DSC-9	4
BSDB32303T	Data Preparation	DSC-10	4
BSDB32304T	Mathematical foundation for DS	SEC-1	4
BSDB32301P	Database Management Systems Lab	CC-3	2
BSDB32302P	Data mining and Visualization Lab	DSC-11	2
BSDB32303P	Data Preparation Lab	DSC-12	2
DABB32305T	Drug Abuse: Problem, Prevention and Management **		0
	Semester IV		
BSDB32401T	Computer System Architecture	CC-4	4
BSDB32402T	Data Communication and Networks	DSC-13	4
BSDB32403T	Machine Learning	DSC-14	4

BSDB32404T	Introduction to Analytics and AI	SEEC-2	4
BSDB32401P	Computer System Architecture Lab	CC-4	2
BSDB32402P	Digital Communication and Networks	DSC-15	2
BSDB32403P	Machine Learning Lab	DSC-16	2
	Total Credits of Semester IV		22
	Semester V		
BSDB33501T	Introduction to Deep Learning	DSE -1	4
BSDB33502T	Big Data Management	DSE-2	4
BSDB33503T	Software Project Management	DSE-3	4
BSDB33504T	Social Network Analysis	SEC-3	4
BSDB33501P	Introduction to Deep Learning Lab	DSE -4	2
BSDB33502P	Big Data Management Lab	DSE-5	2
BSDB33503P	Software Project Management Lab	DSE-6	2
	Semester VI		
BSDB33601T	e-Platforms for ML	SEEC-4	4
BSDB33602T	Applications of Data Science	DSE-6	4
BSDB33603T	Technical Report Writing	DSE-7	4
BSDB33602P	Applications of Data Science Lab	DSE-8	2
BSDB33603P	Technical Report Writing Lab	DSE-9	2
BSDB33604P	Project		6
	Total Credits of the course		132

*__Int__: Continuous Internal Assessment, **Ext**: End Semester Examination, **TOT:** Total, **MAX:** Maximum.

**These are compulsory non-credit qualifying papers. Though they are non-credit courses, it is important for a learner to qualify them failing which the degree will not be awarded.

CC-Core Course, AECC- Ability Enhancement Compulsory Course, SEC- Skill Enhancement Elective Course, DSC- Discipline Specific Elective Course, TSG, DAB and HRD - Compulsory Non Credit Qualifying Course.

G. Procedure for Admissions

Notifications regarding admission will be published in the leading national and regional newspapers. In addition to this, all the required information will be updated regularly on the university website

- **Programme Duration:** 3 years to 6 years
- **The Medium of Examination:** English
- **Eligibility**
  Passed 10+2 in Non-Medical/Commerce/Arts with Mathematics/Statistics/Computer Science/Economics as one of the optional subjects or its equivalent examination stream conducted by Board recognized or established by Central/State Government through a legislation
  OR
  Those candidates who have passed their Matriculation examination AND have also passed three year Diploma in any Trade from Punjab State Board of Technical Education & Industrial Training, Chandigarh or such Examination from any other recognized State Board of Technical Education.

- **Instructional Delivery Mechanisms:** The course has been programmed with the aim to reach the distant and those lacking access to a regular mode of education. The courses' instructions and specially prepared study material will be made available through study centres and digital media like e-mail, website etc. Limited face to face contact sessions will be held at the study centers set up by the university as close as possible to the learner's home. Communication with the university and interaction between the teacher and the learners will be further facilitated using electronic media options like telephone, e-mails, chat sessions, videoconferencing and teleconferencing, if and when required.

  Besides this, Counseling Sessions will be held at all the study centres regularly during weekends. The university will also conduct live/virtual classes for learners using modern ICT methods. However, to ensure learner participation and interaction, online classes will be blended with face to face discussions and meetings with the learners.

H. Evaluation
The learners' progress is measured through the means of continuous evaluation and end semester examinations.

- **Continuous Internal assessment through assignments**

Assignments help the learners to recapitulate the theory and go back to the text again in case they are unable to answer a particular question. Thus, assignments also help to reinforce learning in distance and open learning system of education. The assignments will consist of a set of questions and activities that have to be answered by the programme participants by remaining at their own place.

Two assignments will be submitted for a 3 or 4-credit course and three assignments will be submitted by the learner for a 6-credit course. The assignments will cover all or any types of questions (long answer type, short answer type, objective type, multiple choice questions and case studies).

Learners will be required to obtain thirty percent marks as pass percentage in each assignment separately. Each assignment will carry 100 marks. In the final result, assignments will carry thirty percent weightage.

- **Semester End Examination**

Semester end examination is the major component of the evaluation system and carries seventy percent weightage in the final result. The university will conduct end semester examination twice a year i.e., in June and in December. The learners can take the examination only after the completion of the course, failing which they can take the same in December or June of subsequent years but within the total span of the programme. In case any student fails to get a passing score in the semester end examination, they will be eligible to reappear in the next semester end examination for that course as and when it is held but within the total span of the programme only.

In order to claim Bachelor's Degree in Science (Data Science), the learner is required to score at least 30% marks in both continuous evaluations (i.e.in assignments) as well as in semester end examinations separately. Besides, at least 35% marks in each course is required in the overall computation.

- **Updated Notification for the Learners**

The information regarding the university policies and procedures, academic activities like assignment submissions, question papers, results and other notices related to examination and evaluation will be uploaded on the official website of the university.

I. Laboratory Support

Modernize computer lab at the Learning Centre will be provided with all latest computers and software required for this course.

J. Library Resources

The students may avail the library facilities at their study centres.

K. Cost Estimation

The cost of the programme will be as per the fee decided upon.

L. Quality Assurance Mechanism

The university has constituted a "Centre of Internal Quality Assurance (CIQA) as per UGC (Open and Distance Learning) Regulations, 2017.

M. Programme Outcome

By the end of the course, the programme participants will :

- Enhance their knowledge and skills in both computer science and statistical modeling for data-intensive problem solving and scientific discovery.

- Equip with software engineering and machine learning skills to design and implement efficient, data-driven solutions to real-world problems.

- train students for careers and advanced studies in a wide range of applied computer science, engineering, business, and biotechnology disciplines

- will have a broad foundation of knowledge and skills and cultivate a commitment to life-long learning.

PART II

SYLLABI OF THE COURSES
For
SESSIONS 2021-22, 2022-23, 2023-24

B.Sc. (Data Science)
Core Course (CC)
Semester I
BSDB31101T: Problem Solving using Computers

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

Objective of this paper is to explain the basic of Python concepts objects, data structures and concepts related to Methods and Functions in python. Paper explicate Object Oriented Programming with Python and comprehend the concepts related to Python Generators and file handling.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

<div align="center">

Section A

</div>

Unit I: **Introduction:** Python installation and setup, Command line Basics; Python Objects and Data Structures Basics: Introduction to Python data types, Variable assignments, Numbers, String, String methods, Lists

Unit II: **Python Comparison Operators:** Chaining comparison operators with logical operators, Pass Break and continue.

Unit III: **Program Flow control in Python:** If Elif and Else statements in python, for loops, While loops

Unit IV: **Methods and Functions in python:** Introduction to functions, Def keyword, User defined functions, arguments and parameters, Parameter naming in python

Section B

Unit V: Object Oriented Programming: Introduction, Classes and objects, attributes and methods, Inheritance and polymorphism, Special methods; Modules and Packages: Pip install and PyPi.

Unit VI: Errors and Exception Handling: Introduction to errors, Built-in errors, raising errors in python, Pylint overview

Unit VII: Python Generators: Yielding and Generator function, Making an iterable from a generator, Generator expressions and performance.

Unit VIII: File handling in Python: Files in python, importing own files, Read and writing text files, working with CSV, XML and JSON files.

Suggested Readings
1. Timothy Budd, Exploring Python, TMH, 1st Ed, 2011
3. Allen Downey, Jeffrey Elkner, Chris Meyers , How to think like a computer scientist : learning with Python , Green Tea Pr, 2002
4. Paul Barry, Head First Python: A Brain-Friendly Guide, O′Reilly, 2nd ed. 2016
5.Udemy, https://www.udemy.com/course/complete-python-bootcamp/
6. Udemy, https://www.udemy.com/course/python-the-complete-python-developer-course/

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester I
BSDB31102T: Fundamental of IT

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

This course introduces the concepts of computer basics and working with word processor, spreadsheet, and presentation software packages. Basic concepts of information technology have also been explained in this course.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

<div align="center">Section A</div>

Unit I: Computer Fundamentals: Block diagram of a Computer, Characteristics of Computers, Hardware, Software, Machine Language, Assembly Language and Assembler, High Level Language and Compiler v/s Interpreter. Input Devices: Keyboard, Mouse, Joystick, Track Ball, Touch Screen, Light Pen, Digitizer, Scanners, Speech Recognition Devices, Optical Recognition devices – OMR, OBR, OCR. Output Devices: Monitors, Impact Printers - Dot matrix, Character and Line printer, Non Impact Printers – DeskJet and Laser printing, Plotter.

Unit II: Computer Memory: Representation of information: BIT, BYTE, Memory, Memory size; Units of measurement of storage; Main memory: main memory organization, RAM, ROM,

PROM, EPROM, Computer languages: Machine language, assembly language, higher level language, 4GL. Introduction to Compiler, Interpreter, Assembler, System Software, Application Software. Introduction to Internet, WWW and Web Browsers

Unit III: **Basic of Computer networks;** LAN, WAN; Concept of Internet; Applications of Internet; connecting to internet; ISP; Knowing the Internet; Web Browsing software's, Search Engines; Understanding URL; Domain name; IP Address; Using e-governance website Basics of electronic mail; Getting an email account; Sending and receiving emails; Accessing sent emails; Using Emails;

Unit IV: Word Processing Package: Opening, saving and closing an existing document; renaming and deleting files; Using styles and templates: Introduction to templates and styles; applying, modifying; using a template to create a document, creating a template, editing a template, organizing templates, examples of style use, Changing document views

<div align="center">

Section B

</div>

Unit V: Working with text: select, cut, copy, paste, find and replace, inserting special characters, setting tab stops and indents, Formatting text, formatting paragraphs, Formatting pages: Using layout methods, creating headers and footers, Numbering pages, Changing page margins, Adding comments to a document, Creating a table of contents, Creating indexes and bibliographies, Printing a document, Tracking changes to a document.

Unit VI: Making Small Presentation: Basics of presentation software; Creating Presentation: Entering and Editing Text, Inserting And Deleting Slides in a Presentation, Inserting Word Table or An spreadsheet Worksheet, Adding Clip Art Pictures, Inserting Other Objects, Slide Show: Running a Slide Show, Transition and Slide Timings, Automating a Slide Show,

Unit VII: Using Spreadsheet Statistical functions: SUM, COUNT, AVERAGE, MAX, MIN, MEDIAN, MODE PRODUCT SQRT, STDEV.S, ABS, QUARTILE, PERCENTILE, AVERAGEIF, COUNTA, COUNTBLANK, CORREL, Logical operation IF, SUMIF, AVERAGEIF, COUNTIF,

Unit VIII: Formatting Text: Using RIGHT, LEFT, and MID functions; format text by using UPPER, LOWER, and PROPER functions; format text by using the CONCATENATE function, Generating inference from Data: Pivot Table, Creating Charts, Data Cleaning: Removing duplicate values, Text to Columns,

Suggested Readings

1. Nortorn, P. Introduction to Computers, 7th Edition, 2017
2 .Rajaraman, V., Fundamentals of Computers, PHI, 2014.
3. Larry E. Long and Nancy Long, Computers: Information Technology in Perspective, PHI, 2001
4. Andy Channelle, Beginning OpenOffice 3, Apress, 2009

.

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester I
BSDB31103T: Introduction to Data Science

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

To provide strong foundation for data science and application area related to it and understand the underlying core concepts and emerging technologies in data science. Understand data analysis techniques for applications handling large data.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Data Science-a discipline, Landscape-Data to Data science, Data Growth-issues and challenges, data science process. foundations of data science. Messy data, Anomalies and artefacts in datasets. Cleaning data.

Unit II: Introduction data acquisition, Structured Vs Unsructured data, data preprocessing techniques including data cleaning, selection, integration, transformation and reduction, data mining, interpretation.

Unit III: Representation of data: Special types-acoustic, image, sensor and network data. Problems when handling large data – General techniques for handling large data, Distributing data storage and processing with Frameworks

Unit IV: **Data Science Ethics** – Doing good data science – Owners of the data - Valuing different aspects of privacy - Getting informed consent - The Five Cs – Diversity – Inclusion – Future Trends.

Section B

Unit V: Data Wrangling Combining and Merging Data Sets – Reshaping and Pivoting – Data Transformation – String manipulations – Regular Expressions

Unit VI: Data Aggregation and Group Operations Group By Mechanics – Data Aggregation – GroupWise Operations – Transformations – Pivot Tables – Cross Tabulations – Date and Time data types.

Unit VII: Data Modeling: Basics of Generative modeling and Predictive modeling. Charts-histograms, scatter plots, time series plots etc. Graphs, 3D Visualization and Presentation.

Unit VIII: Applications of Data Science: Business, Insurance, Energy, Health care, Biotechnology, Manufacturing, Utilities, Telecommunication, Travel, Governance, Gaming, Pharmaceuticals, Geospatial analytics and modeling

Suggested Readings
1. Sinan Ozdemir, Principles of Data Science, Packt Publishing, 2016
2. Joel Grus: Data Science from Scratch, O'Reilly, 2016
3. Foster Provost & Tom Fawcett: Data Science for Business O'Reilly, 2013
4. Roger D. Peng & Elizabeth Matsui: The Art of Data Science, Lean Publishing, 2015

B.Sc. (Data Science)
Core Course (CC)
Semester I
BSDB31101P: Problem Solving using Computers Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given practical, other List of Lab Assignments – Session wise will be provided to the students.

1) Find the sum of two numbers. (Python basics)
2) Find the maximum of three numbers in python (Comparison Operators)
3) Print all the prime numbers between two numbers. (Loops)
   Define a function to return the number of vowels in a string. (Methods and functions in python)
5) Delete an item from a tuple (List, tuple, Type conversions)
6) Demonstrate the use of polymorphism by creating different functions for sum. (Polymorphism)
7) Suppose you are given two strings. Compare if the two strings are good or not.
   Strings will be good if they contain same words but the order and case of the words can vary.
      For example:
         "Samsung galaxy s10" and "galaxy SAMSUNG s10" are both good strings.
         "Punjab State Open University Patiala" and "Punjab State Open University Chandigarh" are not good strings. (Concept: Strings)
8) Create Fibonacci series using Generators (Python generators)
9) Write an exception for divisibility of a number by 0. (Exception handling)
10) Print the first 10 lines in text file using python (File handling)
11) Find the words in a string that are greater than k length, k is the user's input.
12) Find the number of pairs in a given list.
    e.g. [10,20,20,10,2,20,7] has 3
13) Take a list of integers. Find the pairs which give the minimum difference
14) Take a string as user's input and print the words from the string that doesn't contain any letter between 0 to 9 between them.
15) Copy content from a file in a computer and paste it into another file.
16) Find out all the numbers from a list in python that has sum k, where k is the user's input.
17) Create a linked list using python
18) Find out the sums of all the "strictly decreasing" numbers from a list of numbers.

19) A strictly decreasing number is the one in which all the numbers from the left to right are decreasing. e.g. 542 and 431 are strictly decreasing but 411 and 424 are not.

20) Rearrange words of a string, according to the sum of ASCII values of the letters of the words, the answer should be a single string of the words.

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester I
BSDB31102P: Fundamental of IT Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

Practical List for Word Processor:

1. Create a telephone directory.
    • The heading should be 16-point Arial Font in bold
    • The rest of the document should use 10-point font size
    • Other headings should use 10-point Courier New Font.
    • The footer should show the page number as well as the date last updated.
2. Design a time-table form for your college.
    • The first line should mention the name of the college in 16-point Arial Font and should be bold.
    • The second line should give the course name/teacher's name and the department in 14-point Arial.
    • Leave a gap of 12-points.
    • The rest of the document should use 10-point Times New Roman font.
    • The footer should contain your specifications as the designer and date of creation.
    3. Create the following one page documents.
     (a) Compose a note inviting friends to a get-together at your house, including a list of things to bring with them.
    (b) Design a certificate in landscape orientation with a border around the document.
  4. Create the following document: A newsletter with a headline and 2 columns in portrait orientation, including at least one image surrounded by text.
  5. Prepare a grocery list having four columns (Serial number, the name of the product, quantity and price) for the month of April,
    • Font specifications for Title (Grocery List): 14-point Arial font in bold and italics.
    • The headings of the columns should be in 12-point and bold.
    • The rest of the document should be in 10-point Times New Roman.
    • Leave a gap of 12-points after the title.
  6. XYZ Publications plans to release a new book designed as per your syllabus. Design the first page of the book as per the given specifications.

(a) The title of the book should appear in bold using 20-point Arial font.

(b) The name of the author and his qualifications should be in the center of the page in 16-point Arial font.

(c) At the bottom of the document should be the name of the publisher and address in 16-point Times New Roman.

(d) The details of the offices of the publisher (only location) should appear in the footer.

7. Create the following one page documents. a) Design a Garage Sale sign. b) Make a sign outlining your rules for your bedroom at home, using a numbered list.

8. Enter the following data into a table given on the next page.

Salesperson	Dolls	Trucks	Puzzles
Hardeep Singh	675	536	464
Amitoj Singh	453	853	684
Devinder Singh	754	464	785
Amritpal Singh	875	785	795
Sharanjit Singh	356	854	785

Add a column Region (values: S, N, N, S, S, S) between the Salesperson and Dolls columns to the given table

Sort your table data by Region and within Region by Salesperson in ascending order:

Practical List for Power Point:

1. Create five Power point slides. Each slide should support different format. In these slides explain areas of applications of IT. Make slide transition time as 10 seconds.

2. Create five Power Point slides to give advantages/disadvantages of computer, application of computers and logical structure of computer.

3. Create five Power Point slides detailing the process of internal assessment. It should be a self-running demo.

Practical List for Spreadsheet:

1. Create a student worksheet containing roll numbers, names and total marks. Open a document in Word and insert the excel worksheet using:-i) Copy/Paste ii) Embedding iii) Linking

2. The term wise marks for APS class of 20 students are stored in 3 separate sheets named term1, term2 and term3. Create 4th worksheet that contains student names and their total and average marks for the entire year. Give proper headings using headers. Make the column

headings bold and italic. The 4th worksheet should contain college name as the first line. Make it bold, italic and center it.

3. Consider the following facts

Full Name	Grade	F Basic pay	HRA	PF	Gross	Vehicle Allowance

HRA is calculated as follows:

Grade	HRA %(of Basic)
1	40%
2	35%
3	30%

Gross = Basic + HRA + VA

Net = Gross –PF

PF is 8% for all Grades

VA is 15000, 10000 and 7000 for Grades 1, 2 and 3.

i) Find max, min and average salary of employees in respective Grade

ii) Count no. of people where VA>HRA

iii) Find out most frequently occurring grade.

iv) Extract records where employee name starts with "A" has HRA>10000

v) Print Grade wise report of all employees with subtotals of net salary and also grand totals. Use subtotal command.

vi) Extract records where Grade is 1 or 2 and salary is between 10000 and 20000 both inclusive.

4. Use Spreadsheet Statistical functions: SUM, COUNT, AVERAGE, MAX, MIN, MEDIAN, MODE PRODUCT SQRT, STDEV.S, ABS, QUARTILE, PERCENTILE, AVERAGEIF, COUNTA, COUNTBLANK, CORREL, Logical operation IF, SUMIF, AVERAGEIF, COUNTIF,

5. Generating inference from Data: Pivot Table, Creating Charts, Data Cleaning: Removing duplicate values, Text to Columns,

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester I
BSDB31103P: Introduction to Data Science Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

1. Practical on data preprocessing techniques and data cleaning.
2. Practical on Data Wrangling Combining and Merging Data Sets
3. Practical on Pivot Tables
4. Practical demonstration on String manipulations and regular expressions
5. Practical on Group Operations
6. Working with date and time data types
7. Practical on visualizations: Charts-histograms, scatter plots, time series plots
8. Practical on Combining and Merging Data Sets
9. Program involving Regular Expressions

AE1B31104T: Effective Communication in English

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

The course is designed to develop in learners the vital communication skills which should be integral to personal, social and professional interactions. Besides, the awareness to use verbal and non-verbal skills appropriately and with confidence will help them excel in the academics and their upcoming professional spheres.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give clear instructions to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES:

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section-A

Unit 1: Understanding Human Communication, Constitutive Processes of Communication, Language as a tool of communication, Barriers to Effective communication, Strategies to Overcome the Barriers

Unit II: Communication: Non-Verbal Communication, Importance of Non-Verbal Communication, Non-Verbal Communication and Cultural influences, Body language

Unit III: **Listening Comprehension:** Difference between Hearing and Listening, Barriers to Listening, Effective Listening Strategies, Listening in Conversational Interaction, Listening to Structured Talks

Unit IV: Speaking Skills: Expressions in different Communicative Functions: Asking

Questions; Making Requests and suggestions; Expressing Greetings, Apologies and Gratitude
Job Interviews, Group Discussions, Presentation Skills

Section-B

Unit V: Reading Comprehension: Introduction, Reading Process, Reading different kinds of Texts, Reading Methods- Using KWL for reading comprehension, SQ3R approach.

Unit VI: Effective Written Communication: Constituents of Effective Writing, Coherence and Cohesion for effective writing, Paragraph Development, Note Making.

Unit VII: Business correspondence I: Letter writing, Resume/CV, E mails for Communication

Unit VIII: Business correspondence-II: Writing Reports, Describing Tables and Charts, Meetings: Agenda and Minutes

Suggested Readings

1. Koneru, Aruna. Professional Communication. Delhi: McGraw. 8th Ed, 2017
2. Mahanand, Anand. English for Academic and Professional Skills. Delhi: McGraw, 2013
3. Rani, D Sudha, TVS Reddy, D Ravi, and AS Jyotsna. A Workbook on English Grammar and Composition. Delhi: McGraw, 2012
4. Rizvi, M. Ashraf. Effective Technical Communication. Delhi: McGraw, 2nd Ed. 2017

Pease, Allan and Barbara Pease. The Definitive Book of Body Language. New Delhi: Manjul Publishing House, 2014

B.Sc. (Data Science)
Compulsory Non-Credit Qualifying Course.
Semester I
TSGB31105T: Teachings of Sikh Gurus

Semester II
B.Sc. (Data Science)
Core Course(CC)
Semester II
BSDB31103P: Operating Systems

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

Understanding basics of operating system viz. system programs, system calls, user mode and kernel mode. Working with CPU scheduling algorithms for specific situation, and analyze the environment leading to deadlock and its rectification. Exploring memory management techniques viz. caching, paging, segmentation, virtual memory, and thrashing.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

SECTION A

UNIT- I: Introduction and System Structures: Computer-System Organization, Computer-System Architecture, Operating-System Structure, Operating-System Operations, Process Management, Memory Management, Storage Management, Protection and Security, Computing

Environments, Operating-System Services, User and Operating-System Interface, System Calls, Types of System Calls, System Programs.

UNIT II: Process Management: Process Concept, Process Scheduling, Operations on Processes, Multi-threaded programming: Multithreading Models, Process Scheduling: Basic Concepts, Scheduling Criteria, and Scheduling Algorithms.

Unit III: Deadlock: System Model, Deadlock Characterization, Methods for Handling Deadlocks, Deadlock Prevention, Deadlock Avoidance, Deadlock Detection, Recovery from Deadlock.

UNIT IV: Memory Management: Basic Hardware, Address Binding, Logical and Physical Address, Dynamic linking and loading, Swapping, Contiguous Memory Allocation, Segmentation, Paging, Demand Paging, Page Replacement algorithms

SECTION B

UNIT V: File Systems: File Concept, Access Methods, Directory and Disk Structure, File-System Structure, File-System Implementation, Directory Implementation, Allocation Methods, Free-Space Management.

UNIT VI: **Introduction to Linux:** Linux's shell, Kernel, Features of Linux, Using file system: Filenames, Introduction to different types of directories: Parent, Subdirectory, Home directory; rules to name a directory, Important directories in Linux File System,

UNIT VII: Linux Commands: cal, date, echo, bc, who, cd, mkdir, rmdir, ls, cat cp, rm, mv, more, gzip, tar, File ownership, file permissions, chmod, Directory permission, change file ownership,

UNIT VIII: Shell Scripting: Creating and Executing Shell Programs, Using variables: Assigning a value to a variable, Accessing the value of a variable, Positional Parameters and other Built-In Shell Variables; Special Characters, Conditional Statements : if Statement, case Statement; Iteration Statements : for Statement, while Statement, until Statement

Suggested Readings
1. A Silberschatz, P.B. Galvin, G. Gagne, Operating Systems Concepts, 8th Edition, John Wiley Publications, 2009
2. A.S. Tanenbaum, Modern Operating Systems, 3rd Edition, Pearson Education, 2014
3. G. Nutt, Operating Systems: A Modern Perspective, 2nd Edition Pearson Education, 2000
4. S. Das, Unix Concepts and Applications, 4th edition, McGraw Hill Education, 2017

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester II
BSDB31202T: Statistical Foundation

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective: This course will enable students to understand the fundamentals of statistics to apply descriptive measures and probability for data analysis. Students will able to infer the concept of correlation and regression for relating two or more related variables and probabilities for various events.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A
Unit I: Origin and development of Statistics, Scope, limitation and misuse of statistics. Types of data: primary, secondary, quantitative and qualitative data. Types of Measurements: nominal, ordinal, discrete and continuous data.

Unit II: Presentation of data by tables: construction of frequency distributions for discrete and continuous data, graphical representation of a frequency distribution by histogram and frequency polygon, cumulative frequency distributions. Classification and Graphical representation of data (Pie Chart, Bar Diagram, Histogram, Frequency Polygon, Ogive Curve, etc.).

Unit III: Measures of Central Tendency – Arithmetic Mean, Median and Mode and its Graphical representation, Measures of dispersion – range, variance, mean deviation, standard deviation and coeff. of variation, Concepts and Measures of Skewness and Kurto.

Unit IV: Descriptive Statistics, Exploratory data analysis, Coefficient of variation, Data **visualization**, Scatter diagram, Grouped data,

Section B

Unit V: Correlation: Scatter plot, Karl Pearson coefficient of correlation, Spearman's rank correlation coefficient, multiple and partial correlations (for 3 variates only). Regression: Introduction to regression analysis: Modelling a response, overview and applications of regression analysis, Simple linear regression (Two variables)

Unit VI: Mathematical and Statistical probability, Elementary events, Sample space, Compound events, Types of events, Random experiment, sample point and sample space, event, algebra of events.

Unit VII: Definition of Probability: classical, empirical and axiomatic approaches to probability, properties of probability. Theorems on probability, conditional probability and independent events

Unit VIII: Statistical inference, Concept of Random Variable, Probability Mass Function & Density Function, Mathematical Expectation (meaning and properties), Moments, Moment Generating Function and Characteristic Function

Suggested Readings
1. Gupta, S.C. and Kapoor, V.K.: Fundamentals of Mathematical Statistics, Sultan & Chand & Sons, New Delhi, 11th Ed,
2. Hastie, Trevor, et al. The elements of Statistical Learning, Springer,
3. Ross, S.M., Introduction to Probability and Statistics, Academic Foundation,
4. Papoulis, A. and Pillai, S.U. , Probability, Random Variables and Stochastic Processes, TMH,

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester II
BSDB31203T: Introduction to Logic

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective: This course will enable students to understand the fundamentals of logic. Students will be able to infer the concept Logic and inference.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten ques
4. tions from this section.
5. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
6. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit- I: Definition, Nature and Significance of Logic; Nature of Implication; Truth and Validity; Laws of Thought, Nature of Proposition; Traditional classification of Propositions; Quality, Quantity and Distribution of Terms.

Unit-II: Immediate Inferences: Conversion, Obversion & Contraposition; Square of Opposition; Mediate Inferences; Categorical Syllogism: Rules, Fallacies & Validity through Venn-Diagrammes.

Unit-III: Truth-functional Logic: Truth-functional compound statements; Negation, Conjunction, Disjunction and Implication. Validity & Invalidity through Truth-table Method; Statement Forms: Tautology, Contradictory and Contingent.

Unit IV: Propositional Logic: Syntax of Propositional Logic, Logical Connectives: Truth Tables, Validity, Consistency, Logical Equivalence. Conjunctive and Disjunctive Normal Forms

Section B

Unit V: **Predicate Logic:** Quantifiers, Translating simple syllogistic sentences to Predicate logic, Semantics of Predicate Logic, Conversion to Clausal form Resolution, Unification, Truth, satisfiability, validity in Predicate Logic.

Unit VI: **Fuzzy Logic**: Basic concepts of fuzzy set theory – operations of fuzzy sets – properties of fuzzy sets – Crisp relations – Fuzzy relational equations – operations on fuzzy relations – fuzzy systems,

Unit VII: **Prolog:** Introduction, Variables and atoms, Facts and predicates, data types, goal finding, Clauses, Central Idea of Prolog, Execution of Prolog Programs, backtracking, simple object, compound objects,

Unit VIII: Arithmetic Operators, Program Termination, Use of cut and fail predicates, Satisfiability: Use Unification, recursion, lists, simple input/output, dynamic database.

Suggested Readings
1. Cohen & Nagal : Introduction to Logic and Scientific Method, Macmillan Publishing Company, London, 1934
2. Copi, Cohen, Jetli : Introduction to Logic, Pearson Education, 12th Edition, 2013
3. Timothy J.Ross, Fuzzy logic with Engineering Applications, 3rd Ed. McGraw Hill, 2011
4. Ivan Bratko, PROLOG Programming For Artificial Intelligence, Addison Wesley, 2011

B.Sc. (Data Science)
Core Course (CC)
Semester II
BSDB31201P: Operating Systems Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given Practicals, List of Lab Assignments – Session wise will be given to the student

1. Installation of Unix/Linux operating system.
2. Study of Unix/Linux general purpose utility command listed in the syllabus.
(cal, date, echo, bc, who, cd, mkdir, rmdir, ls, cat cp, rm, mv, more, gzip, tar, File ownership, file permissions, chmod, Directory permission, change file ownership)
3. Study of Unix/Linux filesystem (tree structure).
4. Write a shell script program to display "HELLO WORLD".
5. Write a shell script program to develop a scientific calculator
6. Write a shell Script program to check whether the given number is even or odd.
7. write a shell script to find the factorial of given integer
8. Write a Shell script that accepts a filename, starting and ending line numbers as arguments and displays all the lines between the given line numbers?

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester II

BSDB31202P: Statistical Foundation Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given practicals List of Lab Assignments – Session wise will be given to the student

1. Performing Data exploration and preprocessing in Python
2. Classification and Graphical representation of data
3. Measurement of of Central Tendencies
4. Measurement of dispersion
5. Implementing correlation and regression analysis
6. Implementing normal and binominal distribution
7. Implementing conditional Probability
8. Drawing random samples using random number tables.
9. Implementing Density function

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester II
BSDB31203P: Introduction to Logic Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given Practical, List of Lab Assignments – Session wise will be given to student

1. Write simple fact for the statements using PROLOG.
2. Write predicates One converts centigrade temperatures to Fahrenheit, the other checks if a temperature is below freezing.
3. Program to add two numbers.
4. Program to categorize animal characteristics.
5. Program to read address of a person using compound variable.
6. Program of fun to show concept of cut operator.
7. Program to count number of elements in a list.
8. Program to reverse the list.
9. Program to append an integer into the list.
10. Program to replace an integer from the list.
11. Program to delete an integer from the list .
12. Program to show concept of list.
13. Program to demonstrate family relationship

B.Sc. (Data Science)
Ability Enhancement Compulsory Course (AECC)
Semester II
AE2B31204T: ENVIRONMENT STUDIES

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective:
The objective of this paper is to create awareness about environmental problems among learners. The paper imparts basic knowledge about the environment and its problems and attempts to motivate learners to participate in improvement.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:
6. The syllabus prescribed should be strictly adhered to.
7. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions each from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
8. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any 10 questions from this section.
9. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
10. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES:
Candidates are required to attempt any two questions each from the sections A, and B of the question paper, and any ten short answer questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

SECTION-A
The multidisciplinary nature of environmental studies. Definition, scope and importance Concept of Biosphere – Lithosphere, Hydrosphere, Atmosphere.
Ecosystem & Biodiversity Conservation
Ecosystem and its components, Types of Ecosystems
Biodiversity - Definition and Value, Threats to biodiversity and its conservation
Level of biological diversity: genetic, species and ecosystem diversity; bio-geographic zones of India; biodiversity patterns and global biodiversity hot spots.
India as Mega-biodiversity nation; Endangered and endemic species of India.

Ecosystem and biodiversity services: Ecological, economic, social, ethical, aesthetic and informational value.

Natural Resources–Renewable And Non Renewable Resources

Land resources and land use change; land degradation, soil erosion and desertification. Deforestation: causes and impacts due to mining, dam building on environment, Forests, Biodiversity and tribal populations.

Water: Use and over-exploitation of surface and ground water, Floods, droughts, conflicts over water (international & inter-state)

Energy resources: renewable and nonrenewable energy sources, use of alternate energy sources, growing energy needs, case studies.

Environmental Pollution

Environmental Pollution : types, causes, effects and controls; Air, Water, Soil and noise pollution. Nuclear hazards and human health risks Solid waste management, Source Segregations : Control measures of urban and Industrial waste. Pollution case studies.

SECTION-B

Environmental Protection Laws In India

Environmental protection act for; Air (Prevention and control of pollution), Water (Prevention and Control of pollution), Wild life, Forest Conservation, Issues involved in the enforcement of environmental legislation.Role of an individual in prevention of pollution.

Environmental policies & Practices; Climate change, global warming, ozone layer depletion, acid rain and imapcts on human communities and agriculture.

Human Communities and the Environment

Human population growth: Impacts on environment, human health and welfare, Sanitation & Hygiene. Resettlement and rehabilitation of project affected persons; case studies. Disaster management: floods, earthquake, cyclones and landslides. Environment movements: Chipko, Silent valley, Bishnois of Rajasthan. Environmental ethics: Role of Indian and other religions and cultures in environmental conservation for a Clean-green pollution free state.

Environmental communication and public awareness, case studies (e.g., CNG vehicles in Delhi)

Road safety awareness

Concept and significance of Road safety, Traffic signs, Traffic rules, Traffic Offences and penalties, How to obtain license, Role of first aid in Road Safety.

Stubble Burning

Meaning of Stubble burning. Impact
on health & environment.

Management and alternative uses of crop stubble.

Environmental Legislations and Policies for Restriction of Agriculture Residue Burning in Punjab.

Suggested Readings

1. Carson, R. Silent Spring, Houghton Mifflin Harcourt, 2002..
2. Gadgil. M., & Guha,R. This Fissured Land : An Ecological History of India. Univ. of California Press, 1993.
3. Gleeson, B. and Low, N.(eds.) Global Ethics and Environment, London, Routledge, 1999.
4. Gleick,P.H. Water in Crisis.Pacific Institute for Studies in Dev. Environment & Security. Stockholam Env. Institute, Oxford Univ. Press, 1993..
5. Groom, Martha J., Gary K. Meffe, and Carl Ronald Carroll. Principles of Conservation Biology. Sunderland : Sinauer Associates, 2006.
6. Grumbine, R. Edward, and Pandit, M.K. Threats from India's Himalays dams. Science,339:36-37,2013.
7. McCully,P. Rivers no more: the environmental effects of dams (pp.29-64). Zed Books, 1996
8. McNeill, John R. Something New Under the Sun : An Environmental History of the Twentieth Century, 2000.
9. Pepper, I.L., Gerba ,C.P & Brusseau, M.L. Environmental and Pollution Sciences. Academic Press, 2011.
10. Rao, M.N. & Datta, A.K. Waste Water Treatment. Oxford and IBH Publishing Co. Pvt.Ltd, 1987..
11. Raven, P.H., Hassenzahl, D.M. & Berg, L.R., Environment. 8Th edition. John Wiles & Sons, 2012.
12. Rosencranz, A., Divan, S., & Nobie, M.L. Environmental law and policy in India. Tripathi, 2001.
13. Sengupta, R. Ecology and economics: An approach to sustainable development. OUP, 2003.
14. Singh, J.S., Singh, S.P. and Gupta, S.R. Ecology, Environmental Science and Conservation. S. Chand Publishing, New Delhi, 2014.
15. Sodhi, N.S. Gibson, L. & Raven, P.H. (eds). Conservation Biology: Voices from the Tropics. John Wiley & Sons, 2013.
16. Wilson, E.O. The Creation: An appeal to save life on earth. New York: Norton, 2006.
17. World commission on Environment and Development. Our Common Future. Oxford University Press, 1987.
   .

B.Sc. (Data Science)
Compulsory Non-Credit Qualifying Course
Semester II
HRDB31205T: Human Rights and Duties

Max marks:100
External Marks:70
Internal Marks:30

Objective: **Pass Percentage35%**

The objective of the course is to impart Learners basic knowledge about human rights as well as duties, and to enable them to meet challenges of human rights violations.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:
11. The syllabus prescribed should be strictly adhered to.
12. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions each from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
13. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any 10 questions from this section.
14. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
15. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES:
Candidates are required to attempt any two questions each from the sections A, and B of the question paper, and any ten short answer questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A
(Introduction to Human Rights)
Unit I- Foundational Aspects: Meaning, Characteristics; Classification; Generations of Human Rights.
Unit II-Constitutional-Legal Recognition in India: Fundamental Rights; Directive Principles of State Policy.

Section B
(Introduction to Human Duties)
Unit III- Conceptual Perspective: Meaning, Nature & Characteristics of Human Duties; Classification of Human Duties; Relevance of Human Duties.
Unit VI- Recognition to Human Duties in India: Fundamental Duties in Indian Constitution Part IV A.

Suggested Readings:

1. United Nations. The United Nations and Human Rights 1945-1995. Geneva: United Nations Blue Books Series, Vol. VII, 1996.
2. Sastry, S. N. Introduction to Human Rights and Duties. Pune: University of Pune Press, 2011.
3. Mertus, Julie. The United Nations and Human Rights-A Guide for a New Era. London: Routledge, 2009.
4. Donnelly, Jack. Universal Human Rights in Theory and Practice. New York: Cornell University Press, 2013.
5. Hammarberg, Thomas. Taking Duties Seriously- Individual Duties in International Humanitarian Law. Versoix: International Council on Human Policy, 1999.
6. Miller P. Frederic, et al. Fundamental Rights, Directive Principles and Fundamental Duties in India. New York: VDM Publishing, 2009.
7. Deol, Satnam Singh. Human Rights in India-Theory and Practice. New Delhi: Serials Publications, 2011.

B.Sc. (Data Science)
Core Course (CC)
Semester III
BSDB32301T: Data Base Management System

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objectives

This course explains fundamental elements of relational database management systems and made student familiar with the basic concepts of relational data model, entity-relationship model, relational database design, relational algebra and SQL

Section A

UNIT I: Introduction to DBMS - Overview of DBMS, Basic DBMS terminology, Data independence. Architecture of a DBMS, Disadvantages of Traditional DBMS, Advantages and Characteristics of DBMS.

UNIT II: Introduction to data models –Relational Keys: Primary Key, Foreign Key, Candidate Key, Super Key etc., and Integrity Constraints, Relational model, Relational schema Hierarchical model, and Network model.

UNIT III: Conceptual data modeling using E-R data model -Entities, attributes, relationships, generalization, specialization, specifying constraints, Conversion of ER Models to Tables, Practical problems based on E-R data model.

UNIT IV: Normal Forms - Functional Dependency, Multi valued dependencies and Joined dependencies, INF, 2NF, 3NF, BCNF, 4NF, 5NF.

Section B

UNIT V: Structured Query Language - Introduction to SQL, data types, DDL, DML, DCL, querying database tables, Data Definition Language (DDL), Creating Tables, Inserting and updating values into a Table.

UNIT VI: Data Manipulation Language: Various form of SELECT- simple, using special operators, aggregate functions, group by clause, sub query, joins, co-related sub query, union clause, exist operator, Aggregate Functions.

UNIT VII: VIEWS - Introduction to views, data independence, Statements on Join Views, Dropping a VIEW. Database security, Security Techniques, Two Phase Locking Techniques.

UNIT VIII: Data Control Operations   - GRANT command, REVOKE command, COMMIT and ROLLBACK. Concurrency Control Techniques, Recovery Control techniques.

Suggested Readings
1. Silverschatz A., Korth F. H. and Sudarshan S., Database System Concepts, Tata McGraw Hill 6th ed., 2019
2. Elmasri R. and Navathe B. S., Fundamentals of Database Systems, Pearson 7th ed, 2016
3. Bayross I., SQL, PL/SQL the Programming Language of Oracle, BPB Publications 4th Ed., 2009

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester III
BSDB32302T: Data Mining and Visualization

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

This course helps student understand the need for Data Mining and advantages to the business world. They will get a clear idea of various classes of Data Mining techniques, their need, scenarios (situations), scope of their applicability and use of data visualization techniques

Section-A

Unit -1 Data Mining: Introduction, Scope, What is Data Mining; How does Data Mining Works, Predictive Modeling: Data Mining and Data Warehousing: Architecture for Data Mining: Profitable Applications: Data Mining Tools

Unit -II: Data Pre-processing: Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.

Unit- III: Data Mining Techniques- An Overview, Data Mining Versus Database Management System, Data Mining Techniques- Association rules, Classification, Regression, Clustering, Neural networks.

Unit -IV Clustering: Introduction, Cluster Analysis, Clustering Methods- K means, Hierarchical clustering, Agglomerative clustering, Divisive clustering, evaluating clusters.

Section-B

Unit -V Applications of Data Mining: Introduction, Business Applications Using Data Mining- Risk management and targeted marketing, Customer profiles and feature construction, Medical applications (diabetic screening), Scientific Applications using Data Mining, Other Applications.

Unit -VI Data Visualization: Introduction, Acquiring and Visualizing Data, Simultaneous acquisition and visualization, Applications of Data Visualization, Keys factors of Data Visualization (Control of Presentation, Faster and Better JavaScript processing, Rise of HTML5, Lowering the implementation Bar)

Unit -VII Exploring the Visual Data Spectrum: charting Primitives (Data Points, Line Charts, Bar Charts, Pie Charts, Area Charts), Exploring advanced Visualizations (Candlestick Charts, Bubble Charts, Surface Charts, Map Charts, Infographics).

Unit -VIII Visualizing data Programmatically, Starting with Google charts (Google Charts API Basics, A Basic bar chart, A basic Pie chart, Working with Chart Animations).

Suggested Readings

1. Jiawei Han, Micheline Kamber and Jian Pei, Data Mining Concepts and Techniques, Third Edition, 2000
2. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Introduction to Data Mining, Pearson 2005,
3. M. Kantardzic, "Data Mining: Concepts, Models, Methods, and Algorithms", 2nd edition, Wiley-IEEE Press, 2011
4. Jon Raasch, Graham Murray, Vadim Ogievetsky, Joseph Lowery, JavaScript and jQuery for Data Analysis and Visualization, 2014
5. Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly, 2007

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester III
BSDB32303T: Data Preparation

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

This course will make student familiar with the methods of data gathering and preparing the data for processing. Student will explore the sampling and data processing techniques.

Section A

Unit I: Data collection and management: Introduction, Sources of data, Data collection and APIs, Exploring and fixing data, Data storage and management, using multiple data sources, Data Quality, Addressing Data Quality Issues.

Unit II: Data Collection: Experiments and surveys, Collection of Primary and Secondary Data, selection of appropriate method for data collection, case study method. Introduction to Data Preparation process, Some problems in preparation process, Missing values and Outliers, types of Analysis.

Unit III Data Preparation: Data formats, parsing and transformation, Scalability and real-time issues. Data Cleaning: Consistency checking, Heterogeneous and missing data, Data Transformation and segmentation

Unit IV: Research Design: Meaning, Need for Research Design, Features of a Good Design, Important Concepts relating to Research Design, Different Research Designs. cluster analysis: Introduction, distance measures Clustering algorithms, agglomerative clustering.

Section B

Unit V: Sampling distributions: Non - central chi - square, t and F distributions and their properties - Distributions of quadratic forms under normality -independence of quadratic form and a linear form - Cochran's theorem.

Unit VI: Simple random sampling (WR and WOR) - Use of Random number Table, Stratified Random Sampling: Properties, Systematic Sampling: Estimation of the Mean and Variance – Comparison of Simple, Stratified and Systematic Sampling

Unit VII Testing f Hypothesis: Hypothesis, Basic Concepts Concerning Testing the Hypotheses, Test Statistic and Critical region, critical value and Decision Rule, Procedure for

Hypothesis Testing, Hypothesis Testing for – Means, Proportions, variance, difference of two mean, difference of two proportions, difference of two variances; P-Value approach, power of test,

Unit VIII: Concepts of time series – Components of time series – Additive and multiplicative models for the analysis of time series - Measurement of trend by (i) Graphic method, (ii) Semi Average method, (iii) Method of Curve Fitting by principle of least squares, (iv)Method of Moving Averages

Suggested Readings

1. C. R. Kothari – Research Methodology Methods and Techniques - New Age International Publishers, 3rd Edition, 2014.
 2. Garg, B.L., Karadia, R., Agarwal, F. and Agarwal, An introduction to Research Methodology, RBSA Publishers. 2002
3. Gupta, S.C. and Kapoor, V.K.: Fundamentals of Applied Statistics, Sultan Chand & Co., 11th ed., 2001
4. William G. Cochran.: Sampling Techniques, John Wiley Sons, 3rd edition, 1977

B.Sc. (Data Science)
Semester III
DABB32305T: Drug Abuse: Problem, Prevention and Management **

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective
The objective of the course is to spread awareness amongst learners regarding social, psychological and physical effects of drug abuse, and familiarize them with the policies and treatment services available.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions each from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any 10 questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
Candidates are required to attempt any two questions each from the sections A, and B of the question paper, and any ten short answer questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section- A
Unit I- Introduction to the Problem: Concept (what constitutes drug abuse); Nature (vulnerable age groups, signs and symptoms); Causes(physiological, psychological, sociological); Consequences (for individuals, families, society and nations)

Unit II- Management of Drug Abuse: Medical Management (Medication for treatment and to reduce withdrawal effects, Drug De-addiction clinics, Relapse management); Psycho-Social Management(Counselling , family and group therapy, behavioural and cognitive therapy, Environmental Intervention).

Section- B

Unit III- Social Efforts for Prevention of Drug Abuse: Role of Family and Social Institutions(Parent child relationship, Family support, Supervision, Shaping values, Active Scrutiny; School Counselling, Teacher as role-model. Parent-Teacher-Health Professional Coordination, Random testing on students; Media: Restraint on advertisements of drugs, advertisements on bad effects of drugs, Publicity and media, Campaigns against drug abuse, Educational and awareness program

Unit IV- Political Efforts for Prevention of Drug Abuse: NDPS Act, Statutory warnings, Policing of Borders, Checking Supply/Smuggling of Drugs, Strict enforcement of laws, Time bound trials.

Suggested Readings

1. Inciardi, J.A., The Drug Crime Connection. Beverly Hills: Sage Publications, 1981.
2. Modi, Ishwar and Modi, Shalini, Drugs: Addiction and Prevention, Jaipur: Rawat Publication, 1997
3. Sain, Bhim Drug Addiction Alcoholism, Smoking obscenity New Delhi: Mittal Publications, 1991.
4. Sandhu, Ranvinder Singh, Drug Addiction in Punjab: A Sociological Study. Amritsar: Guru Nanak Dev University, 2009.
5. Singh, Chandra Paul Alcohol and Dependence among Industrial Workers: Delhi: Shipra, 2000.
6. World Drug Report, United Nations office of Drug and Crime. 8. World Drug Report 2010, United Nations office of Drug and Crime, 2011.
7. The Narcotic Drug and Psychotropic Substances Act, 1985, Universal, Delhi, 2012.

B.Sc. (Data Science)
Skill Enhancement Course (SEC)
Semester III
BSDB32304T: Mathematical Foundation for Data Science

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

This course helps student understand the basic concepts Mathematics that will be helpful to them in data science.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions each from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any 10 questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A, and B of the question paper, and any ten short answer questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Basics of Data Science- Introduction, Importance of data science, statistics and optimization from a data science perspective, structured thinking for solving data science problems using mathematics

Unit II: Linear Equations- Solution of Simultaneous Linear Equations (upto two variable case), Solution of Quadratic Equations. Series: Arithmetic Progression Series, Geometric Progression Series

Unit III: Permutations and Combinations, Binomial Theorem, Determinants with simple applications for solution of linear simultaneous equations using Cramer's Rule.

Unit IV: Matrices with simple application for solution of linear simultaneous equations using matrix inversion method, Eigenvalues and eigenvectors; Matrix factorizations.

Section B

Unit V: Functions. Graphical representations of functions, limits and continuity of functions, first principle of differential calculus, derivations of simple algebraic functions

Unit VI: Applications- Applications of derivatives in Economic and Commerce. Maximum and minimum.

Unit VII: General form of linear Programming, formulating Linear Programming Problems assumptions, Graphic Method.

Unit VIII: The Standard Maximum and Minimum Problems, Simplex Method, Duality, Dual Linear Programming Problems

Suggestive Readings

1. G. Strang, Introduction to Linear Algebra, Wellesley-Cambridge Press, Fifth edition, USA, 2016
2. Bendat, J. S. and A. G. Piersol, Random Data: Analysis and Measurement Procedures. 4th Edition. John Wiley & Sons, Inc., NY, USA, 2010.
3. Cathy O'Neil and Rachel Schutt, Doing Data Science, O'Reilly Media, 2013

B.Sc. (Data Science)
Core Course (CC)
Semester III
BSDB32301P: Data Base Management System Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given practical, List of Lab Assignments – Session wise will be given to student

1. Design a Database and create required tables.
2. Apply the constraints like Primary Key , Foreign key, NOT NULL to the tables.
3. Write a sql statement for implementing ALTER, UPDATE and DELETE
4. Write the queries to implement the joins
5. Write the query for implementing the following functions: MAX(),MIN(),AVG(),COUNT()
6. Write the query to implement the concept of Integrity constrains
7. Write the query to create the views
8. Perform the queries for triggers
9. Perform the following operation for demonstrating the insertion ,updation and deletion
10. using the referential integrity constraints
11. Write the query for creating the users and their role.

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester III
BSDB32302P: Data Mining and Visualization Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given practical, List of Lab Assignments – Session wise will be given to student

1. Explore WEKA Data Mining/Machine Learning Toolkit

- Downloading and/or installation of WEKA data mining toolkit,
- Understand the features of WEKA toolkit such as Explorer, Knowledge Flow interface,
- Experimenter, command-line interface.
- Navigate the options available in the WEKA
- (ex. Select attributes panel, Preprocess panel, Classify panel, Cluster panel, Associate panel
- and Visualize panel)
- Study the arff file format
- Explore the available data sets in WEKA.
- Load a data set (ex. Weather dataset, Iris dataset, etc.)
- Load each dataset and observe the following:

    1. List the attribute names and they types

    2. Number of records in each dataset

    3. Identify the class attribute (if any)

    4. Plot Histogram

    5. Determine the number of records for each class.

    6. Visualize the data in various dimensions

2. Perform data pre-processing tasks and Demonstrate performing association rule mining on data sets.

3. Demonstrate performing classification on data sets

- Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier
output. Compute entropy values, Kappa statistic.

- Extract if-then rules from the decision tree generated by the classifier, Observe the confusion matrix.
- Load each dataset into Weka and perform Naïve-bayes classification and k-Nearest Neighbour classification. Interpret the results obtained.
- Plot RoC Curves
- Compare classification results of ID3, J48, Naïve-Bayes and k-NN classifiers for each dataset, and deduce which classifier is performing best and poor for each dataset and justify.

4. Demonstrate performing clustering of data sets

- Load each dataset into Weka and run simple k-means clustering algorithm with different values of k (number of desired clusters). Study the clusters formed. Observe the sum of squared errors and centroids, and derive insights.
- Explore other clustering techniques available in Weka.
- Explore visualization features of Weka to visualize the clusters. Derive interesting insights and explain.

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester III
BSDB32303P: Data Preparation Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Assignment List of Lab Assignments – Session wise will be given to student

1. Experiment on Collecting Primary and Secondary Data
2. Experiments on parsing and transformation of Data
3. Experiments for data Segmentation
4. Experiments on Clustering algorithms
5. Experiments on Non - central chi – square
6. Experiments on t and F distributions
7. Experiment on Random, Systematic Sampling and Stratified Random Sampling
8. Experiments to test hypothesis.
9. Experiments on time series
    a. Graphic method,
    b. Semi Average method,
    c. Method of Curve Fitting by principle of least squares,
    d. Method of Moving Averages

B.Sc. (Data Science)
Core Course (CC)
Semester IV
BSDB32401T: Computer System Architecture

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

This course will illustrate various elementary concepts of Digital Electronics including Number System, Boolean algebra and Combinational Circuits and comprehend the design of basic computer using instruction formats, addressing modes and various memory management techniques and algorithms.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES:

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Basics of Data Representation- Number System, Conversions of Number Systems, 1's and 2's Complements, fixed and floating point representation, character representation, addition, subtraction, magnitude comparison.

Unit II: **Introduction to Boolean algebra** - Logic gates, Boolean algebra, K-Maps, Sum of Products, Product of Sums.

Unit III: Combinational circuits and Sequential Circuits: decoders, multiplexors, Encoders, DE-multiplexers Half Adders, Full Adders, Flip Flops, registers, counters and memory units.

Unit IV: Basic Computer Organization and Design- Computer Architecture, Structure, Computer registers, Common Bus Systems, Arithmetic, Logical, Shift Microperations, and Design of ALU.

Section B

Unit V: Timing and Control Unit-Instruction cycle, Memory reference instructions, Register reference instructions, Input-output instructions, Design of Timing and Control Unit.

Unit VI: Design of Central Processing Unit: Register organization, stack organization, Register Organization, one address instructions, two address instructions, and three address instructions. Instruction formats, addressing modes.

Unit VII: Input-Output Organization: I/O interfaces, Data transfer schemes. I/O control mechanisms - Program controlled, Interrupt controlled and DMA controller.

Unit VIII: **Memory Unit**: Memory hierarchy, High-speed memories, Organization of a Cache memory unit, Virtual memory, Memory management.

Suggested Readings
1. Mano, Morris M., Computer System Architecture, 3rd ed.,Prentice Hall, 2007
2. Hayes, J.P., Computer Architecture and Organization, McGraw Hill, 1998
3. Hennessy, J.L., Patterson, D.A, and Goldberg, D., Computer Architecture A Quantitative Approach, Pearson Education Asia, 2005
4. Leigh, W.E. and Ali, D.L., System Architecture: software and hardware concepts, South Wester Publishing Co., 2000

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester IV
BSDB32402T: Data Communication and Networks

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objectives

To enable student to understand various aspects of data communications and computer networking systems. To examine and understand network protocols and architectures.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES:

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Basic concepts: Components of data communication, modes of communication, standards and organizations, Network Classification, Network Topologies; Transmission media, network protocol; layered network architecture.

Unit II: Models: Overview of OSI reference model; TCP/IP protocol suite. Physical Layer: Cabling, Network Interface Card, Transmission Media Devices- Repeater, Hub, Bridge, Switch, Router, Gateway; Transmission impairments.

Unit III: Data Link Layer: Framing techniques; Error Control; Flow Control Protocols; Shared media protocols - CSMA/CD and CSMA/CA.

Unit IV: Network Layer: Virtual Circuits and Datagram approach, IP addressing methods – Subnetting; Routing Algorithms (adaptive and non-adaptive)

Section B

Unit V: Transport Layer: Elements of transport protocols – Addressing, Connection establishment and release, Flow control and buffering, Transport services, Transport Layer protocol of TCP and UDP.

Unit VI: Session and Presentation Layer: Session Layer – Design issues, remote procedure call. Presentation Layer – Design issues, Data compression techniques, Cryptography.

Unit VII: Application Layer: Application layer protocols and services – Domain name system, HTTP, E-mail, WWW, telnet, FTP, SMTP.

Unit VIII: Network Security: Common Terms, Firewalls, Virtual Private Networks.

Suggestive Readings

1. B.A. Forouzan: Data Communication and Networking, 4th Edition, Tata McGraw Hill, 2017.
2. A. S. Tanenbaum, Computer Networks, 5th Edition, Pearson, 2011
3. D.E. Comer, Internetworking with TCP/IP, Vol. I, Prentice Hall of India, 2015
4. W. Stalling, Data & Computer Communication, 8th edition, Prentice Hall of India, 2013
5. D. Bertsekas, R. Gallager, Data Networks, 2nd edition, Prentice Hall of India. 1992

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester IV
BSDB32403T: Machine Learning

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objectives

This course aims to provide sound foundation to fundamental concepts of machine learning and its application and prepare students for advanced research and real time problem solving in machine learning and related fields. This course will help students to understand the concepts related to regression, classification, clustering, concept clarity on deep learning, dimensionality reduction, model selection and boosting.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER:

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Introduction: Applications of Machine learning, machine learning as a future; Data Pre-processing: Importing the libraries, Importing the dataset, taking care of missing data, encoding categorial data, Splitting the dataset into training set and test set, Feature scaling.

Unit II: Regression: Simple linear regression, Multiple linear regression, Understanding the P-value, Polynomial regression.

Unit III: Classification: Logistic Regression, K-Nearest Neighbors, Support vector machine, Naïve Bayes, Decision tree classification, Random forest classification.

Unit IV: Clustering: k-means clustering, k means random initialization trap, selecting the number of clusters, Hierarchical clustering.

Section B

Unit V: ANN: Introduction to ANNs, Biological Neural Networks; Usefulness and Applications of ANNs; Architectures of ANNs: Single layer, Multi layer, Competitive layer; Learning: Supervised and Unsupervised; Activation functions; Linear and Non-linear Separability

Unit VI: Supervised Models: Hebb Net: introduction, algorithm, application for AND problem; **Perceptron**: architecture, algorithm, application for OR Problem; ADALINE: architecture, algorithm, application for XOR problem;

Unit VII: **MADALINE:** architecture, algorithm, application for XOR problem; Backpropagation Neural Network: architecture, parameters, algorithm, applications, different issues regarding convergence

Unit VIII Unsupervised Models: Kohonen Self –Organizing Maps: architecture, algorithm, application, Adaptive Resonance Theory: introduction, basic architecture, basic operation, ART1 and ART2

Suggested Readings
1. Andreas C. Müller, Introduction to Machine Learning with Python: A Guide for Data Scientists, 2016 ,Sarah Guido
2. E. Alpaydin, Introduction to Machine Learning, 3rd Edition, PHI Learning, 2015
3. K. P. Murphy, Machine Learning:A Probabilistic Perspective, MIT Press, 2012
4. https://www.udemy.com/course/machinelearning/

B.Sc. (Data Science)
Skill Enhancement Course (SEC)
Semester IV
BSDB32404T: Introduction of Analytics and AI

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

To provide an overview of an exciting growing field of big data analytics and introduce the tools required to manage and analyze big data like Hadoop, No-Sql Map-Reduce. This course will enable students to have skills that will help them to solve complex real-world problems of AI for decision support

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: **Big Data Analytics**: Overview and Evolution of Big data, characteristics Big Data and its importance, Drivers for Big data, Big data analytics, Role of Big data in Industry

Unit II: Big Data Applications: Characteristics of Big Data Applications, Perception and Quantification of Value, Understanding the Big Data Storage.

Unit III: Map Reduce Framework: General Overview of High-Performance Architecture, HDFS, Map-Reduce and YARN – Map Reduce Programming Model.

Unit IV: Introduction to Hadoop: Hadoop Architecture, Apache Hadoop Eco-System, Moving Data in and out of Hadoop, Understanding inputs and outputs of Map Reduce, shuffling, sorting, grouping, Data Serialization**.**

Section B

Unit V: Introduction to AI, Problem Solving, Applications of AI using Big Data, Basic Concepts.

Unit VI: Problem-solving through Search: forward and backward, state-space, blind, heuristic, problem-reduction, neural, and evolutionary search algorithms.

Unit VII: Machine Learning and Knowledge Acquisition: learning from memorization, examples, explanation, and exploration. Learning nearest neighbor, naive Bayes, and decision tree classifiers.

Unit VIII: Analytics and AI Strategy-for Business Transfer Re-engineering Business to think AI and Analytics , Robust Data Monetization Strategy , Accelerated Decision-making with Real-Time Analytics

Suggested Readings
1. Tom White, Hadoop, the Definitive guide‖, O'Reilly Media, 2009
2. Donald Miner, Map Reduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems‖, O'Reilly Media, 2012
3. Nathan Marz, Big Data: Principles and best practices of scalable real-time data systems, Manning Publications, 2015

B.Sc. (Data Science)
Core Course (CC)
Semester IV
BSDB32401P: Computer System Architecture Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

 Besides below given assignment List of Lab Assignments – Session wise will be given to student

1)To study basic gates and verify their truth tables.

2) To design and implement encoder and decoder.

3) To design and implement multiplexer.

4) To design and implement de-multiplexer.

5) To realize Basic gates (AND,OR, NOT) From Universal Gates(NAND & NOR).

6) To study about full adder & verify its truth tab

7) To Design Arithmetic Circuit

8) To test for Shift Circuit

9) To implement addressing modes.

10) To implement instruction cycle.

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester IV
BSDB32402P: Data Communication and Computer Networks Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

1.   Familiarization with transmission media and tools: Coaxial cable, UTP cable, Crimping tool, Connectors etc.
2.   Configuration of TCP/IP Protocols in Windows.
3.   Implementation of resource sharing (file, printer etc.)
4.   Configure DHCP/Static IP for a computer connected to a LAN.
5.   Practical implementation of basic network command and Network configuration commands like ping, ipconfig, netstat, tracert etc. for troubleshooting network related problems.
6.   Familiarization with network simulation tools.

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester IV
BSDB32403P: Machine Learning Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

1)   Write a program to demonstrate the working of decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

2)   Build an Artificial Neural Network by implementing the backpropagation algorithm and test the same using appropriate data set.

3)   Write a program to implement naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

4)   Write a program to construct a Bayesian network considering any medical related dataset at UCI.

5)   Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-means algorithm. Compare the results of these two algorithms and comment the quality of clustering.

6)   Write a program (using Python ML library classes) to implement k-Nearest Neighbor algorithm to classify the iris data set. Print both the correct and wrong predictions.

 7)   Implement the non-parametric Locally weighted regression algorithm in order to fit data points. Select the appropriate data set for your experiment and draw graphs.

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester V
BSDB33501T: Introduction to Deep Learning

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

To enable student to understand the concepts related to tensor flow, explicate the tuning of deep neural networks model and understanding the utility of Keras.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

SECTION A

UNIT I: Introduction: Introduction to perceptron, Testing Regular perceptron

UNIT: II: Reusability in tensor flow: Restoring and training on already trained deep neural networks in tensor flow, importing saved tensor flow DNN classifier model

UNIT III: Building regression and time series models: Building a DNN regressor for non-linear time series in tensor flow, visualizing ML results with mathplotlib

UNIT IV: Building unsupervised learning models: Unsupervised learning and k means clustering with tensor flow

<center>**SECTION B**</center>

UNIT V: Tuning deep neural networks model: Optimization algorithms in tensor flow, Activation functions in tensor flow

UNIT VI: Consuming tensor flow via keras: Installing keras, Building DNN classifier with keras, Storing and restoring a trained neural networks model with keras

Unit VII: CNN: Introduction, Architecture of CNN, Convolution Layer, Pooling Layer, Types of Pooling, Activation Functions,

Unit VIII: Role of RELU function and Softmax function in CNN, Fully connected Layer etc.

Suggested Readings

1. Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning (Adaptive Computation and Machine Learning series), MIT Press 2016
2. François Chollet, Deep Learning with Python, Manning Publications, Ist Ed. 2017
3. Seth Weidman, Deep Learning from Scratch: Building with Python from First, Shroff/O'Reilly1st ed. 2019
4. https://www.udemy.com/course/tensorflow-101-introduction-to-deep-learning/

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester V
BSDB33502T: Big Data Management

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objectives

To enable students to understand the basic concepts of big data Management basics, classification and machine learning. To know about Hadoop and Map-Reduce programming to tackle big data problems

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

UNIT I: Big Data: Introduction, Data Storage and Analysis, Comparison with Other Systems, Rational Database Management System.

UNIT II: Big Data Management: challenges, benefits, Big Data Management Services and vendors.

UNIT III: Classification of Big- Data- Introduction**,** Architecture of Big Data**,** Structured Data, Unstructured Data, Semi Structured Data and their algorithms.

UNIT IV: Processing Big Data- Processing Big data through Acquisition, preprocessing and preparation.

<center>**Section B**</center>

UNIT V: Streaming Data- Streaming data, benefits of streaming data, examples of streaming data, Challenges in working with streaming data, Streaming of data Processing BigData with Apache SPARK.

UNIT VI: No-SQL Databases- Types of NoSQL databases, Advantages of NoSQL, Use of NoSQL, SQL vs NoSQL, Schema-less Models.

UNIT VII: No-SQL Tools- Introduction to MongoDB, its key features, Core Server tools, Creating and Querying through Indexes, Document-Oriented, principles of schema design, constructing queries on Databases.

UNIT VIII: Hbase – Analyzing big data with twitter – Big data for E-Commerce, Big data for blogs – Review of Basic Data Analytic Methods.

Suggested Readings
1. Mike Frampton, Mastering Apache Spark, Packt Publishing, 2015.
2. Mohammed Guller, Big Data Analytics with Spark, Apress, 2016

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester V
BSDB33503T: Software Project Management

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

To understand the fundamental principles of Software Project management. This course will enable students to comprehend the fundamentals of managing and optimizing the software development process , Contract Administration, Costing and Budgeting.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section - A

Unit -I Software Project Management, Project Evaluation and Planning - Activities in Software Project Management, Overview of Project Planning, Stepwise planning, Software processes and process models.

Unit -II Estimation and Budgeting of Projects, Cost Benefit Analysis, Cash Flow Forecasting, Cost-Benefit Evaluation Techniques, Risk Evaluation. Project costing, COCOMO 2, Staffing pattern, Effect of schedule compression.

Unit -III: Project Sequencing and Scheduling Activities, Scheduling resources, Critical path analysis, Network Planning, Risk Management, Managing Risks, Risk Planning and Control.

Unit -IV: Project Monitoring and Control- Collecting Data, Visualizing Progress, Cost Monitoring, review techniques, project termination review, Earned Value analysis.

Section -B

Unit -V: Configuration Management: Software Configuration Management (SCM) – Baselines, SCM Process; Version Control; Change Control; Configuration Audit; Status Reporting; Goals of SCM.

Unit VI: Team Development: Basic Concepts; Organization Types – Centralized-control team organization, Decentralized-control team organization, Mixed-control team organization; Dispersed and Virtual Teams, Communication Plans, Leadership.

Unit -VII: Agile Software Development: Scrum, Extreme Programming , Lean development, Kanban, DevOps. People Management- Introduction, Understanding Behavior, Organizational Behavior, Selecting the Right Person for The Job. Managing knowledge in Agile projects.

Unit -VIII: Software Quality Assurance: Software Quality Assurance Activities; Software Quality Standards – ISO Standard, Capability Maturity Model (CMM), Comparison between ISO 9001 & SEI CMM.

Suggested Readings
1. Zykov SV, Singh A, Agile Enterprise Engineering: Smart Application of Human Factors: Models, Methods, Practices, Case Studies, Springer.
2. Royce, Software Project Management, Pearson Education. 1968
3. Ian Sommerville, Software Engineering, Seventh Edition, Pearson Education. 2005
4. R.S. Pressman, Software Engineering: A Practitioner's Approach, Sixth Edition, Tata McGraw-Hill, 2009

B.Sc. (Data Science)
Skill Enhancement Course (SEC)
Semester V
BSDB33504T: Social Network Analysis

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

This course will elaborate basic concepts and theories of network analysis in the social sciences and examine the ways in which networks can contribute to the explanation of Social, political, economic and cultural phenomena.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit 1: **Introduction** to Social Network Analysis (SNA): definition and origin, core features of the SNA, Foundation of social network analysis.

Unit II: Graph Theory: graph basics, graph representation, Matrix Representation, types of graph, and graph algorithms for Social Network Analysis.

Unit III: **Community detectio**n: Clustering, Community structure, Modularity.

Unit IV: Network models: connected components, giant component, diameter, searching algorithms

Section B

Unit V: Predictive modeling: link/attribute prediction, Influence in Social networks

Unit VI: Sentiment Analysis, Recommendation in Social Networks: Collaborative Filtering, and Content based Recommendation Systems

Unit VII: Social network analysis case studies: Twitter, Facebook.

Unit VIII: Modelling and Visualization, Visualizing online social networks, Practical Applications for Prediction of sentiments, human behavior.

Suggested Readings

1. Ian McCulloh, Helen Armstrong, and Anthony Johnson, Social Network Analysis with applications, WILEY Publisher, 2013
2. Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, Social Media Mining: An Introduction, Arizona State University, 2017
3. Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques: Morgan Kaufmann Publishers, 2013

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester V
BSDB33501P: Introduction to Deep Learning Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

1) Write a program to implement perceptron.
2) Write a program to implement AND, OR gates using perceptron.
3) To implement Crab classification using pattern net.
4) Write a program to implement Wine classification using black propagation.
5) Write a program to implement classification of linearly separable data with a perceptron.
6) To study long short term memory for time series prediction.
7) To study the convolutional neural network and recurrent neural networks.
8) To study the Image Net, Google Net, Resent convolution neural networks.

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester V
BSDB33502P: Big Data Management Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

1.   Collection of Big data

2.   Preprocessing and Cleaning

3.   Data Streaming through SPARK

4.   Installation of Mongo DB

5.   No-SQL Queries Manipulation and creation

6.   Execution for Big data applications like Twitter

7.   HBASE for data analytics

8.   Query Execution on HBASE

9.   Viewing and Querying Temporary Files

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester V
BSDB33503P: Software Project Management Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

1. Identifying the Requirements from Problem Statements.
2. Create Project Plan
   Specify project name and start (or finish) date.
   ▪ Identify and define project tasks.
   ▪ Define duration for each project task.
   ▪ Define milestones in the plan
   ▪ Define dependency between tasks
3. Estimation of Project Metrics: Using Basic COCOMO model to estimate project parameters.
4. Modeling UML Use Case Diagrams and Capturing Use Case Scenarios
5. Use MS Project to create a series of tasks leading to completion of a project of your choice.
   For your project, you need to:
   • Set start or ending dates.
   • Develop a list of tasks that need to be completed.
   • Establish any sub tasks and create links.
   • Execute and Monitor Project Plan
   • Generate Dashboard and Reports

B.Sc. (Data Science)
Skill Enhancement Course (SEC)
Semester VI
BSDB33601T: e-Platforms for the Machine Learning

Total Marks: 100
External Marks: 50
Internal Marks: 50
Credits: 4
Pass Percentage: 35%

Objective: To make student familiar with online platform for exceuting machine learning models. Stduents will improve their skill in machine learning and data science

Instruction

Following are the famous online platforms for executing the machine learning models. This has been recommended to the learner to use at-least one platform and run the assignments given by the university using Python and Deep learning. External Examination will be conducted through quiz/viva-vice or any other relevant method.

Google Colab

Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

Kaggle

Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges.

B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester VI
BSDB33602T: Applications of Data Science

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective
To identify the characteristics of datasets and compare the trivial data and big data for various applications. Studnets will be able to implement machine learning techniques and computing environment that will suitable for the applications under consideration.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I:Social Network Analysis for Big Data- Introduction of Social Network Analysis, Graph Essentials, Graph Basics-Graph, Data Collection and Preprocessing.

Unit II: Big Data Analysis and Predictive Analytics -Evolution of analytic scalability, analytic tools- Predictive Analytics using Supervised – Unsupervised learning, Neural networks, Clustering Methods.

Unit III: Financial Data Analytics: Framework, techniques, and metrics, News events impact market sentiment, relating news analytics to stock returns

Unit IV Financial Time Series Analytics and Their Characteristics, Common Financial Time Series models, Autoregressive models, Markov chain models, and Time series models with leading indicators, Long term forecasting

Section B

Unit V Recommendation In Social Media And Behavior: Analytics, Challenges-Classical Recommendation Algorithms, Recommendation Using Social Context, Behavior Analytics: Individual Behavior, Collective Behavior.

Unit VI Introduction to Healthcare Data Analytics, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems.

Unit VII Natural Language Processing and Data Mining for Clinical Text: Core NLP Components, Information Extraction and Named Entity Recognition, Social Media Analytics for Healthcare: Tracking of Infectious Disease Outbreaks, Readmission risk Prediction.

Unit VIII Retail Analytics: Understanding Customer: Profiling and Segmentation, Modelling Churn. Modelling Lifetime Value, Modelling Risk, Market Basket Analysis.

Suggested Readings
1. Mike Frampton, Mastering Apache Spark, Packt Publishing, 2015
2. Mohammed Guller, Big Data Analytics with Spark, Apress, 2016

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester VI

BSDB33603T: Technical Report Writing and IPR

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 35%

Objective

The course will explain the basic related to writing the technical reports and understanding the concepts related to formatting and structuring the report. This will help students to comprehend the concept of proofreading, proposals and practice.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Introduction: An introduction to writing technical reports, technical sentences formation, using transitions to join sentences, Using tenses for technical writing.

Unit II: Planning and Structuring: Planning the report, identifying reader(s), Voice, Formatting and structuring the report, Sections of a technical report, Minutes of meeting writing.

Unit III: **Drafting report and design issues:** The use of drafts, Illustrations and graphics.

Unit IV: Final edits: Grammar, spelling, readability and writing in plain English: Writing in plain English, Jargon and final layout issues, Spelling, punctuation and Grammar, Padding, Paragraphs, Ambiguity.

Section B

Unit V Proofreading and summaries: Proofreading, summaries, Activities on summaries. Presenting final reports: Printed presentation, Verbal presentation skills, Introduction to proposals and practice.

Unit VI Using word processor: Adding a Table of Contents, Updating the Table of Contents, Deleting the Table of Contents, Adding an Index, Creating an Outline, Adding Comments, Tracking Changes, Viewing Changes, Additions, and Comments, Accepting and Rejecting Changes

Unit VII: Using word processor: Working with Footnotes and Endnotes, Inserting citations and Bibliography, Comparing Documents, Combining Documents, Mark documents final and make them read only., Password protect Microsoft Word documents., Using Macros,

Unit VIII: **Nature of Intellectual Property:** Patents, Designs, Trade and Copyright. Process of Patenting and Development: technological research, innovation, patenting, development. International Scenario: International cooperation on Intellectual Property.

Links:

1. https://www.udemy.com/course/reportwriting/
2. https://www.udemy.com/course/professional-business-english-and-technical-report-writing/
3. https://www.udemy.com/course/betterbusinesswriting/
4. T. Ramappa, Intellectual Property Rights Under WTO, S. Chand Publishers, 2008
5. R. P. Merges, P. S. Menell, Mark A. Lemley, Intellectual Property in New Technological Age 1997

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester VI
BSDB33602P: Applications of Data Science Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Besides below given assignment List of Lab Assignments – Session wise will be given to student

1. Data Collection of Social Networks
2. Data Preprocessing
3. Classification of Data
4. Clustering of Data
5. Financial Data Analytics and prediction
6. Twitter Sentimental Analysis
7. Health Care Analytics
8. Retail Anlytics

- Create any links between major tasks.
- Assign a specific amount time for each task.
- Assign resources for each task.
- Create task information for each item you put into the list.

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester VI

BSDB33603P: Technical Report Writing and IPR Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 35%

Assignment List of Lab Assignments – Session wise will be given to the student.

1. Adding a Table of Contents
2. Updating the Table of Contents
3. Deleting the Table of Contents
4. working with Footnotes and Endnotes
5. Inserting citations and a Bibliography
6. Adding an Index
7. Creating an Outline
8. Adding Comments
9. Tracking Changes
10. Viewing Changes, Additions, and Comments
11. Accepting and Rejecting Changes
12. Comparing Documents
13. Combining Documents
14. 1Mark documents as final and make them read only.
15. Password protect Microsoft Word documents.
16. Using Macros
17. Doing prior art search for Patent

B.Sc. (Data Science)
Semester VI

BSDB33604P: Project Work

Total Marks: 100
Credits: 6
Pass Percentage: 35%

Objective

The objective of the project work is to develop a quality software solution for a Data Science Problem. The main objective of this project course is to provide learners a platform to demonstrate their practical and theoretical skills gained during five semesters of study in B.Sc. Programme.

Project Report evaluation is having a weightage of 75% and viva-voce is having a weightage of 25%.

Topic selected for project work should be complex and large enough to justify as a B.Sc. final semester project. The courses studied by the students during the Data Science Programme provide them the comprehensive background knowledge on diverse subject areas which will be helping students in doing project work. The student will receive Project Guidelines along with their 5th semester course material. Students may also download Project Guidelines from the university Website. Students should strictly follow and adhere to the project guidelines.